

**PSYCHOMETRIC IMPLICATIONS OF CORRECTIONS FOR  
ATTENUATION AND RESTRICTION OF RANGE FOR SELECTION  
VALIDATION RESEARCH**

**CARL CHRISTIAAN THERON**

**DISSERTATION PRESENTED FOR THE DEGREE**



**OF  
DOCTOR OF PHILOSOPHY**

**AT THE**

**UNIVERSITY OF STELLENBOSCH**

**PROMOTER: PROF JCD AUGUSTYN**

**MARCH 1999**

## DECLARATION

I, the undersigned, hereby declare that the work contained in this dissertation is my own original work and that I have not previously in its entirety or in part submitted it at any university for a degree

Date: 05-02-99

## OPSOMMING

Die toestande waaronder keuringsprosedures tipes gevalideer word en die toestande waaronder die prosedure uiteindelik gebruik word, verskil normaalweg tot so 'n mate dat die relevansie van die bevindinge in die gedrang kom. Statistiese korreksies vir die geldigheidskoëffisiënt is algemeen beskikbaar. Die res van die argument in terme waarvan 'n keuringsprosedure ontwikkel en regverdig word kan egter ook verwring word deur dieselfde verskille tussen die toestande waaronder die keuringsprosedure gesimuleer word en dié waaronder die prosedure uiteindelik gebruik word. Relatief min kommer bestaan skynbaar egter ten opsigte van die oordraagbaarheid van die besluitnemingsfunksie wat onder die gesimuleerde toestande ontwikkel is of ten opsigte van die verkreeë beskrywings van nut en billikheid. Hierdie toedrag van sake val ietwat vreemd op. Die eksterne geldigheidprobleme geassosieer met validasie-ontwerpe is redelik goed gedokumenteerd. Dit is dus nie asof die psigometrika-literatuur onbewus is van die probleem wat by die veralgemening van resultate van geldigheidstudies ter sprake is nie. Die besluitnemingsfunksie is waarskynlik die spil waarom die keuringsprosedure draai daarin dat dit die onderliggende prestasie-teorie vergestalt, maar meer belangrik, daarin dat dit die daadwerklike aanvaarding en verwerping van applikante bepaal. Indien statistiese korreksies tot die geldigheidskoëffisiënt beperk word bly die besluitnemingsfunksie onveranderd, alhoewel dit ook moontlik verwring mag word deur dieselfde faktore wat sydigheid in die geldigheidskoëffisiënt te weeg bring. Dieselfde logika geld ook ten opsigte van die evaluasie van die besluitnemingsfunksie in terme van nut en billikheid. Indien slegs die geldigheidskoëffisiënt gekorrigeer word bly die "bottom-line" evaluasie van die keuringsprosedure onveranderd. Prakties gesproke dus, verander niks indien statistiese korreksies tot die geldigheidskoëffisiënt beperk word. Die fundamentele navorsingsdoelstelling is om vas te stel of verskille tussen die toestande waaronder die keuringsprosedure gevalideer word, en die toestande waaronder die prosedure uiteindelik gebruik word, sydigheid te weeg bring in die maatstawwe wat vereis word om die keuringsprosedure te spesifiseer en te regverdig; om toepaslike statistiese korreksies vir die geldigheidskoëffisiënt, besluitnemingsreël en beskrywings van nut en billikheid af te lei ten einde die kontekste van simulاسie/validasie en toepassing te versoen; en om vas te stel of sodanige korreksies wel in validasie-navorsing toegepas behoort te word. Die studie verskaf geen ongekwalfiseerde antwoord op die vraag of korreksies vir die verskeie vorms van variانسie-inperking en/of kriterium onbetroubaarheid op die geldigheidskoëffisiënt, die standaardfout van die geldigheidskoëffisiënt of die parameters van die regressie van die kriterium op die voorspeller toegepas behoort te word nie. Die korreksies affekteer wel besluite aangaande die geldigheid van prestasie-hipoteses onder spesifieke toestande. Die korreksies het ook onder bepaalde toestande 'n effek op besluite aangaande applikante deur hul effek op die regressiekoëffisiënte en/of die standaardskattingsfout.

## ABSTRACT

The conditions under which selection procedures are typically validated and those prevailing at the eventual use of a selection procedure normally differ to a sufficient extent to challenge the relevance of the validation research evidence. Statistical corrections to the validity coefficient are generally available. The remainder of the argument in terms of which a selection procedure is developed and justified could, however, also be biased by any discrepancy between the conditions under which the selection procedure is simulated and those prevailing at the eventual use of the selection procedure. Relatively little concern, however, seem to exist for the transportability of the decision function derived from the selection simulation or the descriptions/assessments of selection decision utility and fairness. This seems to be a somewhat strange state of affairs. The external validity problems with validation designs are reasonably well documented. It is thus not as if the psychometric literature is unaware of the problem of generalizing validation study research findings to the eventual area of application. The decision function is probably the pivot of the selection procedure in that it firstly captures the underlying performance theory, but more importantly from a practical perspective, because it guides the actual accept and reject choices of applicants. Restricting the statistical corrections to the validity coefficient would leave the decision function unaltered even though it might also be distorted by the same factors affecting the validity coefficient. Basically the same logic also applies to the evaluation of the decision rule in terms of selection utility and fairness. Correcting only the validity coefficient would leave the "bottom-line" evaluation of the selection procedure unaltered. Restricting the statistical corrections to the validity coefficient basically means that practically speaking nothing really changes. The fundamental research objective is to determine whether any discrepancy between the conditions under which the selection procedure is simulated and those prevailing at the eventual use of the selection procedure produces bias in estimates required to specify and justify the procedure; to delineate appropriate statistical corrections of the validity coefficient, decision rule and descriptions/assessments of selection decision utility and fairness, required to align the contexts of evaluation/validation and application; and to determine whether the corrections should be applied in validation research. The study provides no unqualified answers to the question whether corrections for various forms of range restriction and/or criterion unreliability should be applied to the validity coefficient, the standard error of the validity coefficient or the parameters of the regression of the criterion on the predictor. Under specific conditions the corrections do affect decisions on the validity of performance hypotheses due to its effect on decisions on the significance of the uncorrected versus the corrected validity coefficient. Under specific conditions the corrections do affect decisions on applicants, especially when selection decisions are not restricted by selection quotas, due to its effect on the slope and intercept parameters of the regression of Y on X, and/or due to its effect on the standard error of estimate.



## ACKNOWLEDGEMENTS

It takes more than one person to complete a doctoral thesis and nowhere is it more true than in my case.

I consequently would like to thank the following people and institutions who have made valuable and much appreciated contributions to the completion of this research.

Ψ

My supervisor, Prof JCD Augustyn, for his support, guidance and encouragement and for teaching me what true loyalty to colleagues, friends and family really means.

Ψ

My wife, Antionette and my daughters Michelle and Anelle, for their love, support and encouragement.

Ψ

Prof J Schepers for his invaluable advice, encouragement and recommendations.

Ψ

Prof JM du Toit who inculcated my love and appreciation for psychometrics.

Ψ

My parents for their love, support and investment in me.

Ψ

My colleagues, Coen, Johan, Gawie & Henry who were prepared to take over my lecturing responsibilities so as to enable me to take the sabbatical to complete this research.

Ψ

Gawie Cillié for his willingness to sacrifice long hours in strategic consultation and for his valuable advice and support.

Ψ

Colleagues from the Department of Statistics for their concern, support and advice.

Ψ

The internal and external examiners for pointing out deficiencies in the original manuscript and for their recommendations and advice.

Ψ

The financial assistance of the Centre for Science Development towards this research is hereby acknowledged. Opinions expressed in this thesis and conclusions arrived at, are those of the author and are not necessarily to be attributed to the Centre for Science Development.

TABLE OF CONTENTS

	PAGE
DECLARATION	ii
OPSOMMING	iii
ABSTRACT	iv
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS	vi
LIST OF FIGURES	xviii
LIST OF TABLES	xxxii
1. INTRODUCTION, RESEARCH OBJECTIVE AND OVERVIEW OF THE STUDY	
1.1 Introduction	2
1.2 Personnel Selection	6
1.2.1 An Operational Perspective On Personnel Selection	8
1.2.2 A Strategic Perspective On Personnel Selection	19
1.2.3 An Equity Perspective On Personnel Selection	25
1.3. Summary And Research Objectives	28
1.4. Structural Outline Of Dissertation	31
2. MEASUREMENT THEORY	
2.1 Constructs	35
2.2 The Nature Of Measurement	37
2.3. Measurement Theory	41
2.4 Notation	42
2.5 Classical Measurement Theory	43
2.5.1 Reliability	47

2.5.1.1	Reliability Estimates	50
2.5.1.1.1	Test-Retest Method	54
2.5.1.1.2	Internal Analysis Method	55
2.5.1.1.3	Parallel Forms Method	94
2.5.2	Validity	95
2.6	Summary	104
3.	<b>SEPARATE AND COMBINED CORRECTIONS TO THE VALIDITY COEFFICIENT FOR THE ATTENUATING EFFECT OF THE UNRELIABILITY OF MEASUREMENTS AND RESTRICTION OF RANGE</b>	
3.1.	The Need For Corrections To The Validity Coefficient For The Attenuating Effect Of The Unreliability Of Measurements And Restriction Of Range	106
3.2	The Validity Coefficient Corrected For The Attenuating Effect Of The Unreliability Of The Criterion Measurements	113
3.3	The Validity Coefficient Corrected For The Attenuating Effect Of Explicit Or Implicit Restriction Of Range	117
3.3.1	Univariate Selection	117
3.3.1.1	The Effect Of Case 1 [Case B] Restriction Of Range On The Validity Coefficient	120
3.3.1.2	The Effect Of Case 2 [Case A] Restriction Of Range On The Validity Coefficient	123
3.3.1.3	The Effect Of Case 3 [Case C] Restriction Of Range On The Validity Coefficient	125
3.3.2	Multivariate Selection	138
3.3.3	Accuracy Of Corrections To The Validity Coefficient For The Attenuating Effect Of Explicit Or Implicit Restriction Of Range	138
3.4	The Validity Coefficient Corrected For The Joint Attenuating Effect Of Explicit Or Implicit Range Restriction And The Unreliability Of The Criterion Measurements	143
3.5	Summary	156

<b>4.</b>	<b>STATISTICAL INFERENCE REGARDING CORRELATIONS CORRECTED FOR THE SEPARATE AND COMBINED ATTENUATING EFFECT OF RESTRICTION OF RANGE AND CRITERION UNRELIABILITY</b>	
4.1	Inferential Statistics In Validation Research	157
4.2.	The Need For Information On The Sampling Distribution Of The Validity Coefficient Corrected For The Attenuating Effect Of The Unreliability Of Measurements And/or Restriction Of Range	159
4.3	Different Procedures For Investigating The Sampling Distributions Of Statistical Estimators	160
4.4	Non-Distribution Free Estimates Of Statistical Error	160
4.4.1	Delta Method	161
4.5	Distribution Free Estimates Of Statistical Error	162
4.5.1	Bootstrap Estimates Of Statistical Error	162
4.5.1.1	Bootstrap Confidence Intervals	163
4.5.1.2	Bootstrap Hypothesis Testing	166
4.5.1.3	Bootstrap Sample Size Requirements	167
4.5.2	Jackknife Estimates Of Statistical Error	167
4.6	Expressions For The Standard Error Of The Pearson Correlation Coefficient	168
4.6.1	Obtained, Uncorrected First-Order Pearson Correlation Coefficient	168
4.6.2	Pearson Correlation Coefficient Corrected For Criterion And Predictor Unreliability	170
4.6.3	Pearson Correlation Coefficient Corrected For Criterion Unreliability Only	173
4.6.4	Pearson Correlation Coefficient Corrected For Case 1 [Case B] Restriction Of Range	186
4.6.5	Pearson Correlation Coefficient Corrected For Case 2 [Case A] Restriction Of Range	186
4.6.6	Pearson Correlation Coefficient Corrected For Case C Restriction Of Range	192
4.6.7	Pearson Correlation Coefficient Double Corrected For Restriction Of Range And Criterion Unreliability	201
4.7	Summary	214

5.	EFFECT OF STATISTICAL CORRECTIONS OF THE VALIDITY COEFFICIENT ON DECISIONS ON THE NULL HYPOTHESES	
5.1	The Logic Of [Fisherian] Statistical Significance Testing	215
5.2	Effect Of Corrections For Attenuation And/or Range Restriction To The Pearson Correlation Coefficient On Decisions On Statistical Null Hypotheses	217
5.3	Effect Of Corrections For Attenuation And/or Range Restriction To The Pearson Correlation Coefficient On The Empirically Derived Exceedence Probabilities $\alpha_b$ [Or Achieved Significance Level, ASL]	224
5.3.1	Effect Of Corrections For Attenuation To The Pearson Correlation Coefficient On The Empirically Derived Exceedence Probabilities $\alpha_b$	224
5.3.1.1	The Reliability Coefficient Given A Priori By Theoretical Assumption Or Previously Accepted Knowledge	224
5.3.1.2	The Reliability Coefficient Obtained From An Independent Data Set	226
5.3.1.3	The Reliability Coefficient Obtained From The Same Data Set As $\rho^*[X,Y]$	230
5.3.2	Effect Of Corrections For Restriction Of Range To The Pearson Correlation Coefficient On The Empirically Derived Exceedence Probabilities $\alpha_b$	234
5.3.2.1	Case 2 [Case A] Restriction Of Range	234
5.3.2.2	Case 3 [Case C] Restriction Of Range	237
5.3.3	Effect Of Corrections For Criterion Unreliability And Range Restriction To The Pearson Correlation Coefficient On The Empirically Derived Exceedence Probabilities $\alpha_b$	241
5.4	Effect Of Corrections For Attenuation And/or Range Restriction To The Pearson Correlation Coefficient On The A Priori Probabilities $[1-\beta]$ Or Power Of The Tests Of $H_0$	246



5.4.1	Effect Of Corrections For Attenuation To The Pearson Correlation Coefficient On The A Priori Probabilities $[1-\beta]$ Or Power Of The Tests Of $H_0$	247
5.4.1.1	The Reliability Coefficient Given A Priori By Theoretical Assumption Or Previously Accepted Knowledge	247
5.4.1.2	The Reliability Coefficient Obtained From An Independent Sample	248
5.4.1.3	The Reliability Obtained From The Same Sample As $\rho^*[X,Y]$	249
5.4.2	Effect Of Corrections For Restriction Of Range To The Pearson Correlation Coefficient On The A Priori Probabilities $[1-\beta]$ Or Power Of The Tests Of $H_0$	250
5.4.2.1	Case 2 [Case A] Restriction Of Range	250
5.4.2.2	Case C Restriction Of Range	252
5.4.3	Effect Of Double Corrections For Attenuation And Restriction Of Range To The Pearson Correlation Coefficient On The A Priori Probabilities $[1-\beta]$ Or Power Of The Tests Of $H_0$	256
5.5	Summary	256
6.	<b>EFFECT OF STATISTICAL CORRECTIONS TO THE VALIDITY COEFFICIENT ON THE PARAMETERS OF THE DECISION FUNCTION</b>	
6.1	The Decision Function In Personnel Selection	259
6.2	Effect Of Statistical Corrections For Range Restriction And/or Random Measurement Error In The Criterion On The Decision Function	267
6.2.1	Effect Of Statistical Corrections For Random Measurement Error On The Decision Function	268
6.2.1.1	Effect Of Statistical Corrections For Random Measurement Error In The Criterion On The Decision Function	268

6.2.1.2	Effect Of Statistical Corrections For Random Measurement Error In Both The Criterion And The Predictor On The Decision Function	275
6.2.2	Effect Of Statistical Corrections For Restriction Of Range On The Decision Function	286
6.2.2.1	Case 1 [Case B] Restriction Of Range	286
6.2.2.2	Case 2 [Case A] Restriction Of Range	293
6.2.2.3	Case 3 [Case C] Restriction Of Range	293
6.2.3	Effect Of The Joint Correction For Case 2 [Case A] Restriction Of Range And Criterion Unreliability On The Decision Function	310
6.3	Concluding Comments	315
6.4	Summary	315

## 7. CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER RESEARCH

7.1	Conclusions	316
7.1.1	Reviewing The Necessity And Significance Of The Research	316
7.1.2	Research Objectives	320
7.1.3	Methodology	321
7.1.4	Summary Of Results	322
7.1.4.1	The Effect Of Separate And Combined Statistical Corrections For Restriction Of Range And Random Measurement Error On The Magnitude Of The Validity Coefficient	322
7.1.4.1.1	The Effect Of The Correction For Criterion Unreliability Only On The Magnitude Of The Validity Coefficient	322
7.1.4.1.2	The Effect Of The Correction Case 1 [Case B] Restriction Of Range On The Magnitude Of The Validity Coefficient	322
7.1.4.1.3	The Effect Of The Correction For Case 2 [Case A] Restriction Of Range On The Magnitude Of The Validity Coefficient	323

7.1.4.1.4	The Effect Of The Correction For Case C Restriction Of Range On The Magnitude Of The Validity Coefficient	323
7.1.4.1.5	The Effect Of The Joint Correction For Case 2 [Case A] Restriction Of Range On X And Criterion Unreliability On The Magnitude Of The Validity Coefficient	324
7.1.4.2	The Effect Of Separate And Combined Statistical Corrections For Restriction Of Range And Random Measurement Error On The Magnitude Of The Standard Error Of The Pearson Correlation Coefficient	325
7.1.4.2.1	The Effect Of The Correction For Criterion Unreliability Only On The Magnitude Of The Standard Error Of The Pearson Correlation Coefficient	325
7.1.4.2.2	The Effect Of The Correction For Case 1 [Case B] Restriction Of Range On The Magnitude Of The Standard Error Of The Pearson Correlation Coefficient	327
7.1.4.2.3	The Effect Of The Correction For Case 2 [Case A] Restriction Of Range On The Magnitude Of The Standard Error Of The Pearson Correlation Coefficient	327
7.1.4.2.4	The Effect Of The Correction For Case 3 [Case C] Restriction Of Range On The Magnitude Of The Standard Error Of The Pearson Correlation Coefficient	328
7.1.4.2.5	The Effect Of The Joint Correction For Case 2 [Case A] Restriction Of Range On X And Criterion Unreliability On The Magnitude Of The Standard Error Of The Pearson Correlation Coefficient	329
7.1.4.3	The Effect Of Separate And Combined Statistical Corrections For Restriction Of Range And Random Measurement Error On The Magnitude Of The Empirically Derived Exceedence Probability $\alpha_b$	331

7.1.4.3.1	The Effect Of The Correction For Criterion Unreliability Only On The Magnitude Of The Empirically Derived Exceedence Probability	331
7.1.4.3.1.1	The Reliability Coefficient Given A Priori By Theoretical Assumption Or Previously Accepted Knowledge	331
7.1.4.3.1.2	The Reliability Coefficient Obtained From An Independent Data Set	331
7.1.4.3.1.3	The Reliability Coefficient Obtained From The Same Data Set	332
7.1.4.3.2	The Effect Of The Correction For Case 1 [Case B] Restriction Of Range On The Magnitude Of The Empirically Derived Exceedence Probability	332
7.1.4.3.3	The Effect Of The Correction For Case 2 [Case A] Restriction Of Range On The Magnitude Of The Empirically Derived Exceedence Probability	332
7.1.4.3.4	The Effect Of The Correction For Case 3[i] [Case C[i]] Restriction Of Range On The Magnitude Of The Empirically Derived Exceedence Probability	333
7.1.4.3.5	The Effect Of The Joint Correction For Case 2 [Case A] Restriction Of Range On X And Criterion Unreliability On The Magnitude Of The Empirically Derived Exceedence Probability	334
7.1.4.4	The Effect Of Separate And Combined Statistical Corrections For Restrictions Of Range And Random Measurement Error On The Magnitude Of The A Priori Probability Of Rejecting $H_0$ If $H_0$ Is False	335

7.1.4.4.1	The Effect Of The Correction For Criterion Unreliability Only On The Magnitude Of The A Priori Probability Of Rejecting $H_0$ If $H_0$ Is False	335
7.1.4.4.1.1	The Reliability Coefficient Given A Priori By Theoretical Assumption Or Previously Accepted Knowledge	335
7.1.4.4.1.2	The Reliability Coefficient Obtained From An Independent Sample	335
7.1.4.4.1.3	The Reliability Coefficient Obtained From The Same Sample	335
7.1.4.4.2	The Effect Of The Correction For Case 1 [Case B] Restriction Of Range On The Magnitude Of The A Priori Probability Of Rejecting $H_0$ If $H_0$ Is False	336
7.1.4.4.3	The Effect Of The Correction For Case 2 [Case A] Restriction Of Range On The Magnitude Of The A Priori Probability Of Rejecting $H_0$ If $H_0$ Is False	336
7.1.4.4.4	The Effect Of The Correction For Case 3[i] [Case C[i]] Restriction Of Range On The A Priori Probability Of Rejecting $H_0$ If $H_0$ Is False	336
7.1.4.4.5	The Effect Of The Joint Correction For Case 2 [Case A] Restriction Of Range On X And Criterion Unreliability On The Magnitude Of The A Priori Probability Of Rejecting $H_0$ If $H_0$ Is False	337
7.1.4.5	The Effect Of Separate And Combined Statistical Corrections For Restriction Of Range And Random Measurement Error On The Magnitude Of The Intercept And Slope Parameters Of The Regression Of The Criterion On The Predictor	337



7.1.4.5.1	The Effect Of The Correction For Criterion Unreliability Only On The Magnitude Of The Intercept And Slope Parameters Of The Regression Of The Criterion On The Predictor	337
7.1.4.5.2	The Effect Of Case 1 [Case B] Restriction Of Range On The Magnitude Of The Intercept And Slope Parameters Of The Regression Of The Criterion On The Predictor	338
7.1.4.5.3	The Effect Of The Correction For Case 2 [Case A] Restriction Of Range On The Magnitude Of The Intercept And Slope Parameters Of The Regression Of The Criterion On The Predictor	339
7.1.4.5.4	The Effect Of The Correction For Case 3[i] [Case C[i]] Restriction Of Range On The Magnitude Of The Intercept And Slope Parameters Of The Regression Of The Criterion On The Predictor	340
7.1.4.5.5	The Effect Of The Joint Correction For Case 2 [Case A] Restriction Of Range On X And Criterion Unreliability On The Magnitude Of The Intercept And Slope Parameters Of The Regression Of The Criterion On The Predictor	341
7.1.4.6	The Effect Of Separate And Combined Statistical Corrections For Restriction Of Range And Random Measurement Error On The Magnitude Of The Standard Error Of Estimate Of The Regression Of The Criterion On The Predictor	342
7.1.4.6.1	The Effect Of The Correction For Criterion Unreliability Only On The Magnitude Of The Standard Error Of Estimate Of The Regression Of The Criterion On The Predictor	342

7.1.4.6.2	The Effect Of The Correction For Case 1 [Case B] Restriction Of Range On The Magnitude Of The Standard Error Of Estimate Of The Regression Of The Criterion On The Predictor	343
7.1.4.6.3	The Effect Of The Correction For Case 2 [Case A] Restriction Of Range On The Magnitude Of The Standard Error Of Estimate Of The Regression Of The Criterion On The Predictor	344
7.1.4.6.4	The Effect Of The Correction For Case 3[i] [Case C[i]] Restriction Of Range On The Magnitude Of The Standard Error Of Estimate Of The Regression Of The Criterion On The Predictor	345
7.1.4.6.5	The Effect Of The Joint Correction For Case Restriction Of Range On X And Criterion Unreliability On The Magnitude Of The Magnitude Of The Standard Error Of Estimate Of The Regression Of The Criterion O The Predictor	346
7.1.5	Synopsis Of Findings	347
7.2	Recommendations For Further Research	348
7.2.1	The Effect Of Statistical Corrections For Restriction Of Range And/or Criterion Unreliability On Utility Assessments	348
7.2.1.1	Payoff Defined In Terms Of The Validity Coefficient	350
7.2.1.2	Payoff Defined In Terms Of The Success Ratio	351
7.2.1.3	Payoff Defined In Terms Of The Expected Standardized Criterion Score	353
7.2.1.4	Payoff Defined In Terms Of A Monetary Valued Criterion	355
7.2.1.5	Refining Payoff Defined In Terms Of A Monetary Valued Criterion	359
7.2.1.6	Analysis Of Selection Utility In Terms Of Risk	367

7.2.2 The Effect Of Corrections For Random Measurement Error And Nonrandom Selection On Fairness Assessments Of Selection Procedures	369
7.2.2.1 The Cleary Model Of Selection Fairness	370
7.2.2.2 The Einhorn-Bass Model Of Selection Fairness	371
7.2.2.3 The Thorndike Model Of Selection Fairness	372
7.2.3 Additional Specific Recommendations For Further Research	372
REFERENCES	374
APPENDIX A	385

## LIST OF FIGURES

	PAGE
Figure 2.1	The five pivotal linkages in the interpretive argument underlying selection instrument score interpretation in the context of personnel selection
	101
Figure 3.1	The partially disattenuated validity coefficient as a function of $\rho[X,Y]$ and $\rho_{tt}Y$
	118
Figure 3.2	The partially disattenuated validity coefficient as a function of $\rho[X,Y]$ and $\rho_{tt}Y$
	118
Figure 3.3	The validity coefficient corrected for Case 1 [Case B] restriction of range as a function of K and $\rho[x,y]$
	122
Figure 3.4	The validity coefficient corrected for Case 1 [Case B] restriction of range as a function of K and $\rho[x,y]$
	122
Figure 3.5	The validity coefficient corrected for Case 2 [Case A] restriction of range as a function of K and $\rho[x,y]$
	126
Figure 3.6	The validity coefficient corrected for Case 2 [Case A] restriction of range as a function of K and $\rho[x,y]$
	126
Figure 3.7	The validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of K, $\rho[y,z]$ and $\rho[x,y]$ for $\rho[x,z]$ fixed at 0.2
	131
Figure 3.8	The validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of K, $\rho[y,z]$ and $\rho[x,y]$ for $\rho[x,z]$ fixed at 0.2
	131
Figure 3.9	The validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of K, $\rho[y,z]$ and $\rho[x,y]$ for $\rho[x,z]$ fixed at 0.5
	132
Figure 3.10	The validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of K, $\rho[y,z]$ and $\rho[x,y]$ for $\rho[x,z]$ fixed at 0.5
	132

Figure 3.11	The validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of $K$ , $\rho[y,z]$ and $\rho[x,y]$ for $\rho[x,z]$ fixed at 0.8	133
Figure 3.12	The validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of $K$ , $\rho[y,z]$ and $\rho[x,y]$ for $\rho[x,z]$ fixed at 0.8	133
Figure 3.13	The validity coefficient jointly corrected for Case 2 [Case A] restriction of range and criterion unreliability as a function of $\rho_{tty}$ and $\rho[x,y]$ for $K$ fixed at 1	154
Figure 3.14	The validity coefficient jointly corrected for Case 2 [Case A] restriction of range and criterion unreliability as a function of $\rho_{tty}$ and $\rho[x,y]$ for $K$ fixed at 2	154
Figure 3.15	The validity coefficient jointly corrected for Case 2 [Case A] restriction of range and criterion unreliability as a function of $\rho_{tty}$ and $\rho[x,y]$ for $K$ fixed at 4	155
Figure 3.16	The validity coefficient jointly corrected for Case 2 [Case A] restriction of range and criterion unreliability as a function of $\rho_{tty}$ and $\rho[x,y]$ for $K$ fixed at 5	155
Figure 4.1	The standard error of the partially disattenuated validity coefficient as a function of $\rho[X,Y]$ and $\rho_{ttY}$ for $n$ fixed at 10	174
Figure 4.2	The standard error of the partially disattenuated validity coefficient as a function of $\rho[X,Y]$ and $\rho_{ttY}$ for $n$ fixed at 90	174
Figure 4.3	$V = \sigma[\rho^{\wedge}_c]/\sigma[\rho^{\wedge}]$ as a function of $\rho[X,Y]$ and $\rho_{ttY}$ for $n$ fixed at 10	175
Figure 4.4	$V = \sigma[\rho^{\wedge}_c]/\sigma[\rho^{\wedge}]$ as a function of $\rho[X,Y]$ and $\rho_{ttY}$ for $n$ fixed at 90	175
Figure 4.5	The standard error of the partially disattenuated validity coefficient calculated through Equation 4.23 as a function of $\rho[X,Y]$ and $\rho_{ttY}$ for $n$ fixed at 10	177
Figure 4.6	The standard error of the partially disattenuated validity coefficient calculated through Equation 4.23 as a function of $\rho[X,Y]$ and $\rho_{ttY}$ for $n$ fixed at 90	177



Figure 4.7	The standard error of the partially disattenuated validity coefficient calculated through Equation 4.24 as a function of $\rho[X,Y]$ and $\rho_{tt}Y$ for $n$ fixed at 10	178
Figure 4.8	The standard error of the partially disattenuated validity coefficient calculated through Equation 4.24 as a function of $\rho[X,Y]$ and $\rho_{tt}Y$ for $n$ fixed at 90	178
Figure 4.9	The standard error of the partially disattenuated validity coefficient calculated through Equation 4.25 as a function of $\rho[X,Y]$ and $\rho_{tt}Y$ for $n$ fixed at 10	179
Figure 4.10	The standard error of the partially disattenuated validity coefficient calculated through Equation 4.25 as a function of $\rho[X,Y]$ and $\rho_{tt}Y$ for $n$ fixed at 90	179
Figure 4.11	$V = \sigma[\rho^{\wedge}_{\cdot}] / \sigma[\rho^{\wedge}]$ as a function of $\rho[X,Y]$ , $\rho_{tt}Y$ and $n$ when the standard error of the partially disattenuated validity coefficient is calculated through Equation 4.23 for $n$ fixed at 10	181
Figure 4.12	$V = \sigma[\rho^{\wedge}_{\cdot}] / \sigma[\rho^{\wedge}]$ as a function of $\rho[X,Y]$ , $\rho_{tt}Y$ and $n$ when the standard error of the partially disattenuated validity coefficient is calculated through Equation 4.23 for $n$ fixed at 90	181
Figure 4.13	$V = \sigma[\rho^{\wedge}_{\cdot}] / \sigma[\rho^{\wedge}]$ as a function of $\rho[X,Y]$ , $\rho_{tt}Y$ and $n$ when the standard error of the partially disattenuated validity coefficient is calculated through Equation 4.24 for $n$ fixed at 10	182
Figure 4.14	$V = \sigma[\rho^{\wedge}_{\cdot}] / \sigma[\rho^{\wedge}]$ as a function of $\rho[X,Y]$ , $\rho_{tt}Y$ and $n$ when the standard error of the partially disattenuated validity coefficient is calculated through Equation 4.24 for $n$ fixed at 90	182
Figure 4.15	$V = \sigma[\rho^{\wedge}_{\cdot}] / \sigma[\rho^{\wedge}]$ as a function of $\rho[X,Y]$ , $\rho_{tt}Y$ and $n$ when the standard error of the partially disattenuated validity coefficient is calculated through Equation 4.25 for $n$ fixed at 10	183

Figure 4.16	$V = \sigma[\rho^{\wedge}_c]/\sigma[\rho^{\wedge}]$ as a function of $\rho[X,Y]$ , $\rho_{tt}Y$ and $n$ when the standard error of the partially disattenuated validity coefficient is calculated through Equation 4.25 for $n$ fixed at 90	183
Figure 4.17	The standard error of the validity coefficient corrected for Case 2 [Case A] restriction of range as a function of $K$ and $\rho[x,y]$ for $n$ fixed at 10	188
Figure 4.18	The standard error of the validity coefficient corrected for Case 2 [Case A] restriction of range as a function of $K$ and $\rho[x,y]$ for $n$ fixed at 90	188
Figure 4.19	The standard error of the validity coefficient corrected for Case 2 [Case A] restriction of range as a function of $K$ and $\rho[x,y]$ [rotated $60^\circ$ ] for $n$ fixed at 10	189
Figure 4.20	The standard error of the validity coefficient corrected for Case 2 [Case A] restriction of range as a function of $K$ and $\rho[x,y]$ [rotated $60^\circ$ ] for $n$ fixed at 90	189
Figure 4.21	$V = \sigma[\rho^{\wedge}_c]/\sigma[\rho^{\wedge}]$ as a function of $K$ and $\rho[x,y]$ for $n$ fixed at 10	190
Figure 4.22	$V = \sigma[\rho^{\wedge}_c]/\sigma[\rho^{\wedge}]$ as a function of $K$ and $\rho[x,y]$ for $n$ fixed at 90	190
Figure 4.23	$V = \sigma[\rho^{\wedge}_c]/\sigma[\rho^{\wedge}]$ as a function of $K$ and $\rho[x,y]$ [rotated through $60^\circ$ ] for $n$ fixed at 10	191
Figure 4.24	$V = \sigma[\rho^{\wedge}_c]/\sigma[\rho^{\wedge}]$ as a function of $K$ and $\rho[x,y]$ [rotated through $60^\circ$ ] for $n$ fixed at 90	191
Figure 4.25	The standard error of the validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of $\rho[x,y]$ and $\rho[x,z]$ for $n$ fixed at 10, $K$ fixed at 2 and $\rho[y,z]$ fixed at 0.15	194
Figure 4.26	The standard error of the validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of $\rho[x,y]$ and $\rho[x,z]$ for $n$ fixed at 90, $K$ fixed at 2 and $\rho[y,z]$ fixed at 0.15	194
Figure 4.27	The standard error of the validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of $\rho[x,y]$	

	and $\rho[x,z]$ for $n$ fixed at 10, $K$ fixed at 5 and $\rho[y,z]$ fixed at 0.15	195
Figure 4.28	The standard error of the validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of $\rho[x,y]$ and $\rho[x,z]$ for $n$ fixed at 90, $K$ fixed at 2 and $\rho[y,z]$ fixed at 0.15	195
Figure 4.29	The standard error of the validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of $\rho[x,y]$ and $\rho[x,z]$ for $n$ fixed at 10, $K$ fixed at 2 and $\rho[y,z]$ fixed at 0.25	196
Figure 4.30	The standard error of the validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of $\rho[x,y]$ and $\rho[x,z]$ for $n$ fixed at 90, $K$ fixed at 2 and $\rho[y,z]$ fixed at 0.25	196
Figure 4.31	The standard error of the validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of $\rho[x,y]$ and $\rho[x,z]$ for $n$ fixed at 10, $K$ fixed at 5 and $\rho[y,z]$ fixed at 0.25	197
Figure 4.32	The standard error of the validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of $\rho[x,y]$ and $\rho[x,z]$ for $n$ fixed at 90, $K$ fixed at 5 and $\rho[y,z]$ fixed at 0.25	197
Figure 4.33	The standard error of the validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of $\rho[x,y]$ and $\rho[x,z]$ for $n$ fixed at 10, $K$ fixed at 2 and $\rho[y,z]$ fixed at 0.65	198
Figure 4.34	The standard error of the validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of $\rho[x,y]$ and $\rho[x,z]$ for $n$ fixed at 90, $K$ fixed at 2 and $\rho[y,z]$ fixed at 0.65	198
Figure 4.35	The standard error of the validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of $\rho[x,y]$ and $\rho[x,z]$ for $n$ fixed at 10, $K$ fixed at 5 and $\rho[y,z]$ fixed at 0.65	199

Figure 4.36	The standard error of the validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of $\rho[x,y]$ and $\rho[x,z]$ for $n$ fixed at 90, $K$ fixed at 5 and $\rho[y,z]$ fixed at 0.65	199
Figure 4.37	The standard error of the validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of $\rho[x,y]$ and $\rho[x,z]$ [rotated through $60^\circ$ ] for $n$ fixed at 10, $K$ fixed at 5 and $\rho[y,z]$ fixed at 0.65	200
Figure 4.38	The standard error of the validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of $\rho[x,y]$ and $\rho[x,z]$ [rotated through $60^\circ$ ] for $n$ fixed at 90, $K$ fixed at 5 and $\rho[y,z]$ fixed at 0.65	200
Figure 4.39	$V = \sigma[\rho^{\wedge}_c]/\sigma[\rho]$ as a function of $\rho[x,y]$ and $\rho[x,z]$ for Case 3[i] [Case C[i]] restriction of range; $n$ is fixed at 10, $K$ is fixed at 2 and $\rho[y,z]$ is fixed at 0.15	202
Figure 4.40	$V = \sigma[\rho^{\wedge}_c]/\sigma[\rho]$ as a function of $\rho[x,y]$ and $\rho[x,z]$ for Case 3[i] [Case C[i]] restriction of range; $n$ is fixed at 90, $K$ is fixed at 2 and $\rho[y,z]$ is fixed at 0.15	202
Figure 4.41	$V = \sigma[\rho^{\wedge}_c]/\sigma[\rho]$ as a function of $\rho[x,y]$ and $\rho[x,z]$ for Case 3[i] [Case C[i]] restriction of range; $n$ is fixed at 10, $K$ is fixed at 5 and $\rho[y,z]$ is fixed at 0.15	203
Figure 4.42	$V = \sigma[\rho^{\wedge}_c]/\sigma[\rho]$ as a function of $\rho[x,y]$ and $\rho[x,z]$ for Case 3[i] [Case C[i]] restriction of range; $n$ is fixed at 90, $K$ is fixed at 5 and $\rho[y,z]$ is fixed at 0.15	203
Figure 4.43	$V = \sigma[\rho^{\wedge}_c]/\sigma[\rho]$ as a function of $\rho[x,y]$ and $\rho[x,z]$ for Case 3[i] [Case C[i]] restriction of range; $n$ is fixed at 10, $K$ is fixed at 2 and $\rho[y,z]$ is fixed at 0.35	204
Figure 4.44	$V = \sigma[\rho^{\wedge}_c]/\sigma[\rho]$ as function of $\rho[x,y]$ and $\rho[x,z]$ for Case 3[i] [Case C[i]] restriction of range; $n$ is fixed at 90, $K$ is fixed at 2 and $\rho[y,z]$ is fixed at 0.35	204
Figure 4.45	$V = \sigma[\rho^{\wedge}_c]/\sigma[\rho]$ as a function of $\rho[x,y]$ and $\rho[x,z]$ for Case 3[i] [Case C[i]] restriction of range; $n$ is fixed at 10, $K$ is fixed at 5 and $\rho[y,z]$ is fixed at 0.35	205

Figure 4.46	$V = \sigma[\rho^{\wedge}_c]/\sigma[\rho]$ as function of $\rho[x,y]$ and $\rho[x,z]$ for Case 3[i] [Case C[i]] restriction of range; n is fixed at 90, K is fixed at 5 and $\rho[y,z]$ is fixed at 0.35	205
Figure 4.47	$V = \sigma[\rho^{\wedge}_c]/\sigma[\rho]$ as a function of $\rho[x,y]$ and $\rho[x,z]$ Case 3[i] [Case C[i]] restriction of range; n is fixed at 10, K is fixed at 2 and $\rho[y,z]$ is fixed at 0.65	206
Figure 4.48	$V = \sigma[\rho^{\wedge}_c]/\sigma[\rho]$ as a function of $\rho[x,y]$ and $\rho[x,z]$ for Case 3[i] [Case C[i]] restriction of range; n is fixed at 90, K is fixed at 2 and $\rho[y,z]$ is fixed at 0.65	206
Figure 4.49	$V = \sigma[\rho^{\wedge}_c]/\sigma[\rho]$ as a function of $\rho[x,y]$ and $\rho[x,z]$ for Case 3[i] [Case C[i]] restriction of range; n is fixed at 10, K is fixed at 5 and $\rho[y,z]$ is fixed at 0.65	207
Figure 4.50	$V = \sigma[\rho^{\wedge}_c]/\sigma[\rho]$ as a function of $\rho[x,y]$ and $\rho[x,z]$ for Case 3[i] [Case C[i]] restriction of range ; n is fixed at 90, K is fixed at 5 and $\rho[y,z]$ is fixed at 0.65	207
Figure 4.51	The standard error of the validity coefficient jointly corrected for Case 2 [Case A] restriction of range and criterion unreliability as a function of $\rho[x,y]$ and $\rho_{ttY}$ for n fixed at 10 and K fixed at 2	209
Figure 4.52	The standard error of the validity coefficient jointly corrected for Case 2 [Case A] restriction of range and criterion unreliability as a function of $\rho[x,y]$ and $\rho_{ttY}$ for n fixed at 90 and K fixed at 2	209
Figure 4.53	The standard error of the validity coefficient jointly corrected for Case 2 [Case A] restriction of range and criterion unreliability as a function of $\rho[x,y]$ and $\rho_{ttY}$ for n fixed at 10 and K fixed at 5	210
Figure 4.54	The standard error of the validity coefficient jointly corrected for Case 2 [Case A] restriction of range and criterion unreliability as a function of $\rho[x,y]$ and $\rho_{ttY}$ for n fixed at 90 and K fixed at 5	210
Figure 4.55	The standard error of the validity coefficient jointly corrected for Case 2 [Case A] restriction of range and criterion	



	unreliability as a function of $\rho[x,y]$ and $\rho_{tt}Y$ for $n$ fixed at 10 and $K$ fixed at 2 [rotated through $60^\circ$ ]	211
Figure 4.56	The standard error of the validity coefficient jointly corrected for Case 2 [Case A] restriction of range and criterion unreliability as a function of $\rho[x,y]$ and $\rho_{tt}Y$ for $n$ fixed at 90 and $K$ fixed at 2 [rotated through $60^\circ$ ]	211
Figure 4.57	The standard error ratio $V = \sigma[\rho^*_c]/\sigma[\rho^*]$ as a function of $\rho[x,y]$ and $\rho_{tt}Y$ for $n$ fixed at 10 and $K$ fixed at 2	212
Figure 4.58	The standard error ratio $V = \sigma[\rho^*_c]/\sigma[\rho^*]$ as a function of $\rho[x,y]$ and $\rho_{tt}Y$ for $n$ fixed at 90 and $K$ fixed at 2	212
Figure 4.59	The standard error ratio $V = \sigma[\rho^*_c]/\sigma[\rho^*]$ as function of $\rho[x,y]$ and $\rho_{tt}Y$ for $n$ fixed at 10 and $K$ fixed at 5	213
Figure 4.60	The standard error ratio $V = \sigma[\rho^*_c]/\sigma[\rho^*]$ as a function of $\rho[x,y]$ and $\rho_{tt}Y$ for $n$ fixed at 90 and $K$ fixed at 5	213
Figure 5.1	Probabilities associated with the possible hypothesis test outcomes [directional alternative hypothesis]	217
Figure 5.2	Effect of corrections to the correlation coefficient on statistical power when assuming a constant effect size [graph A represents the situation for the uncorrected correlation and graph B for the corrected correlation]	221
Figure 5.3	Effect of corrections to the correlation coefficient on statistical power when assuming an adjusted effect size [graph A represents the situation for the uncorrected correlation and graph B for the corrected correlation]	223
Figure 5.4	$G = (\rho^*_c/\sigma[\rho^*_c]) / (\rho^*/\sigma[\rho^*]) = Z_c/Z$ as a function of $\rho_{tt}Y$ , the pre-correction correlation coefficient and sample size $n$ when the reliability coefficient $\rho_{tt}Y$ is given a priori by theoretical assumption or previously accepted knowledge; $n$ is fixed at 90	225
Figure 5.5	$G = (\rho^*_c/\sigma[\rho^*_c]) / (\rho^*/\sigma[\rho^*]) = Z_c/Z$ as a function of $\rho_{tt}Y$ , the pre-correction correlation coefficient and sample size $n$ when the reliability coefficient $\rho_{tt}Y$ is given a priori by	

theoretical assumption or previously accepted knowledge;  $n$  is fixed at 120

225

Figure 5.6

$G = (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}]) = Z_c / Z$  as a function of  $\rho_{ttY}$ , the pre-correction correlation coefficient and sample size  $n$  when the reliability coefficient  $\rho_{ttY}$  is obtained from an independent data set;  $n$  is fixed at 10

227

Figure 5.7

$G = (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}]) = Z_c / Z$  as a function of  $\rho_{ttY}$ , the pre-correction correlation coefficient and sample size  $n$  when the reliability coefficient  $\rho_{ttY}$  is obtained from an independent data set;  $n$  is fixed at 20

227

Figure 5.8

$G = (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}]) = Z_c / Z$  as a function of  $\rho_{ttY}$ , the pre-correction correlation coefficient and sample size  $n$  when the reliability coefficient  $\rho_{ttY}$  is obtained from an independent data set;  $n$  is fixed at 90

228

Figure 5.9

$G = (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}]) = Z_c / Z$  as a function of  $\rho_{ttY}$ , the pre-correction correlation coefficient and sample size  $n$  when the reliability coefficient  $\rho_{ttY}$  is obtained from an independent data set;  $n$  is fixed at 120

228

Figure 5.10

$G = (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}]) = Z_c / Z$  as a function of  $\rho_{ttY}$ , the pre-correction correlation coefficient and sample size  $n$  when the reliability coefficient  $\rho_{ttY}$  is obtained from an independent data set [rotated];  $n$  is fixed at 10

229

Figure 5.11

$G = (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}]) = Z_c / Z$  as a function of  $\rho_{ttY}$ , the pre-correction correlation coefficient and sample size  $n$  when the reliability coefficient  $\rho_{ttY}$  is obtained from an independent data set [rotated];  $n$  is fixed at 60

229

Figure 5.12

$G = (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}]) = Z_c / Z$  as a function of  $\rho_{ttY}$ , the pre-correction correlation coefficient and sample size  $n$  when the reliability coefficient  $\rho_{ttY}$  is obtained from the same data set;  $n$  is fixed at

231

Figure 5.13

$G = (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}]) = Z_c / Z$  as a function of  $\rho_{ttY}$ , the pre-correction correlation coefficient and sample size  $n$

	when the reliability coefficient $\rho_{ttY}$ is obtained from an independent data set; n is fixed at 20	231
Figure 5.14	$G = (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}]) = Z_c / Z$ as a function of $\rho_{ttY}$ , the pre-correction correlation coefficient and sample size n when the reliability coefficient $\rho_{ttY}$ is obtained from an independent data set; n is fixed at 90	232
Figure 5.15	$G = (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}]) = Z_c / Z$ as a function of $\rho_{ttY}$ , the pre-correction correlation coefficient and sample size n when the reliability coefficient $\rho_{ttY}$ is obtained from an independent data set; n is fixed at 120	232
Figure 5.16	$G = (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}]) = Z_c / Z$ as a function of $\rho_{ttY}$ , the pre-correction correlation coefficient and sample size n when the reliability coefficient $\rho_{ttY}$ is obtained from an independent data set [rotated]; n is fixed at 20	233
Figure 5.17	$G = (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}]) = Z_c / Z$ as a function of $\rho_{ttY}$ , the pre-correction correlation coefficient and sample size n when the reliability coefficient $\rho_{ttY}$ is obtained from an independent data set [rotated]; n is fixed at 90	233
Figure 5.18	$G = (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}]) = Z_c / Z$ under Case 2 [Case A] restriction of range as a function of $\rho_{ttY}$ and the pre-correction correlation coefficient for n fixed at 10	235
Figure 5.19	$G = (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}]) = Z_c / Z$ under Case 2 [Case A] restriction of range as a function of $\rho_{ttY}$ and the pre-correction correlation coefficient for n fixed at 20	235
Figure 5.20	$G = (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}]) = Z_c / Z$ under Case 2 [Case A] restriction of range as a function of $\rho_{ttY}$ and the pre-correction correlation coefficient for n fixed at 90	236
Figure 5.21	$G = (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}]) = Z_c / Z$ under Case 2 [Case A] restriction of range, as a function of $\rho_{ttY}$ and the pre-correction correlation coefficient for n fixed at 120	236

Figure 5.22	$G = (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}]) = Z_c / Z$ under Case 3[i] [Case C[i]] restriction of range as a function of $\rho^{\wedge}[x,y]$ and $\rho^{\wedge}[x,z]$ for n fixed at 10, K fixed at 2 and $\rho^{\wedge}[y,z]$ fixed at 0.25	238
Figure 5.23	$G = (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}]) = Z_c / Z$ under Case 3[i] [Case C[i]] restriction of range as a function of $\rho^{\wedge}[x,y]$ and $\rho^{\wedge}[x,z]$ for n fixed at 10, K fixed at 4 and $\rho^{\wedge}[y,z]$ fixed at 0.25	238
Figure 5.24	$G = (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}]) = Z_c / Z$ under Case 3[i] [Case C[i]] restriction of range as a function of $\rho^{\wedge}[x,y]$ and $\rho^{\wedge}[x,z]$ for n fixed at 90, K fixed at 2 and $\rho^{\wedge}[y,z]$ fixed at 0.25	239
Figure 5.25	$G = (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}]) = Z_c / Z$ under Case 3[i] [Case C[i]] restriction of range as a function of $\rho^{\wedge}[x,y]$ and $\rho^{\wedge}[x,z]$ for n fixed at 90, K fixed at 2 and $\rho^{\wedge}[y,z]$ fixed at 0.25	239
Figure 5.26	$G = (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}]) = Z_c / Z$ under Case 3[i] [Case C[i]] restriction of range as a function of $\rho^{\wedge}[x,y]$ and $\rho^{\wedge}[x,z]$ for n fixed at 60, K fixed at 4 and $\rho^{\wedge}[y,z]$ fixed at 0.10	240
Figure 5.27	$G = (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}]) = Z_c / Z$ under Case 3[i] [Case C[i]] restriction of range as a function of $\rho^{\wedge}[x,y]$ and $\rho^{\wedge}[x,z]$ for n fixed at 60, K fixed at 4 and $\rho^{\wedge}[y,z]$ fixed at 0.75	240
Figure 5.28	$G = (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}])$ as a function of $\rho^{\wedge}[x,y]$ and $\rho_{ttY}$ when correcting the obtained correlation coefficient for criterion unreliability and [Case 2 [Case A]] restriction of range simultaneously; n is fixed at 10 and K is fixed at 1	242
Figure 5.29	$G = (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}])$ as a function of $\rho^{\wedge}[x,y]$ and $\rho_{ttY}$ when correcting the obtained correlation coefficient for criterion unreliability and [Case 2 [Case A]] restriction of range simultaneously; n is fixed at 10 and K is fixed at 2	242
Figure 5.30	$G = (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}])$ as a function of $\rho^{\wedge}[x,y]$ and $\rho_{ttY}$ when correcting the obtained correlation coefficient for criterion unreliability and [Case 2 [Case A]] restriction of range simultaneously; n is fixed at 10 and K is fixed at 3	243

- Figure 5.31  $G = (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}])$  as a function of  $\rho^{\wedge}[x,y]$  and  $\rho_{ttY}$  when correcting the obtained correlation coefficient for criterion unreliability and [Case 2 [Case A]] restriction of range simultaneously; n is fixed at 10 and K is fixed at 4 243
- Figure 5.32  $G = (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}])$  as a function of  $\rho^{\wedge}[x,y]$  and  $\rho_{ttY}$  when correcting the obtained correlation coefficient for criterion unreliability and [Case 2 [Case A]] restriction of range simultaneously; n is fixed at 120 and K is fixed at 1 244
- Figure 5.33  $G = (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}])$  as a function of  $\rho^{\wedge}[x,y]$  and  $\rho_{ttY}$  when correcting the obtained correlation coefficient for criterion unreliability and [Case 2 [Case A]] restriction of range simultaneously; n is fixed at 120 and K is fixed at 2 244
- Figure 5.34  $G = (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}])$  as a function of  $\rho^{\wedge}[x,y]$  and  $\rho_{ttY}$  when correcting the obtained correlation coefficient for criterion unreliability and [Case 2 [Case A]] restriction of range simultaneously; n is fixed at 120 and K is fixed at 3 245
- Figure 5.35  $G = (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}])$  as a function of  $\rho^{\wedge}[x,y]$  and  $\rho_{ttY}$  when correcting the obtained correlation coefficient for criterion unreliability and [Case 2 [Case A]] restriction of range simultaneously; n is fixed at 120, K is fixed at 4 245
- Figure 5.36  $J = (\rho^{\wedge 2} / \sigma[\rho^{\wedge}]) / (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c])$  as a function of  $\rho^{\wedge}[X,Y]$  and  $\rho^{\wedge}_{ttY}$  for n fixed at 10 251
- Figure 5.37  $J = (\rho^{\wedge 2} / \sigma[\rho^{\wedge}]) / (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c])$  as a function of  $\rho^{\wedge}[X,Y]$  and  $\rho^{\wedge}_{ttY}$  for n fixed at 90 251
- Figure 5.38  $J = (\rho^{\wedge} / \sigma[\rho^{\wedge}]) / (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c])$  under Case 3[i] [Case C[i]] restriction of range as a function of  $\rho^{\wedge}[x,z]$  and  $\rho^{\wedge}[x,y]$  for n fixed at 10, K fixed at 4 and  $\rho^{\wedge}[y,z]$  fixed at 0.25 253
- Figure 5.39  $J = (\rho^{\wedge} / \sigma[\rho^{\wedge}]) / (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c])$  under Case 3[i] [Case C[i]] restriction of range as a function of  $\rho^{\wedge}[x,z]$  and  $\rho^{\wedge}[x,y]$  for n fixed at 90, K fixed at 4 and  $\rho^{\wedge}[y,z]$  fixed at 0.25 253

Figure 5.40	$J = (\rho^{\wedge}/\sigma[\rho^{\wedge}]) / (\rho^{\wedge}_c/\sigma[\rho^{\wedge}_c])$ under Case 3[i] [Case C[i]] restriction of range as a function of $\rho^{\wedge}[x,z]$ and $\rho^{\wedge}[x,y]$ for n fixed at 90, K fixed at 2 and $\rho^{\wedge}[y,z]$ fixed at 0.15	254
Figure 5.41	$J = (\rho^{\wedge}/\sigma[\rho^{\wedge}]) / (\rho^{\wedge}_c/\sigma[\rho^{\wedge}_c])$ under Case 3[i] [Case C[i]] restriction of range as a function of $\rho^{\wedge}[x,z]$ and $\rho^{\wedge}[x,y]$ for n fixed at 90, K fixed at 2 and $\rho^{\wedge}[y,z]$ fixed at 0.65	254
Figure 5.42	$J = (\rho^{\wedge}/\sigma[\rho^{\wedge}]) / (\rho^{\wedge}_c/\sigma[\rho^{\wedge}_c])$ under Case 3[i] [Case C[i]] restriction of range as a function of $\rho^{\wedge}[x,z]$ and $\rho^{\wedge}[x,y]$ for n fixed at 90, K fixed at 4 and $\rho^{\wedge}[y,z]$ fixed at 0.15	255
Figure 5.43	$J = (\rho^{\wedge}/\sigma[\rho^{\wedge}]) / (\rho^{\wedge}_c/\sigma[\rho^{\wedge}_c])$ under Case 3[i] [Case C[i]] restriction of range as a function of $\rho^{\wedge}[x,z]$ and $\rho^{\wedge}[x,y]$ for n fixed at 90, K fixed at 4 and $\rho^{\wedge}[y,z]$ fixed at 0.65	255
Figure 5.44	$J = (\rho^{\wedge}/\sigma[\rho^{\wedge}]) / (\rho^{\wedge}_c/\sigma[\rho^{\wedge}_c])$ under the joint correction for Case 2 [Case A] restriction of range and criterion unreliability as a function of the initial uncorrected effect size estimate and $\rho_{ttY}$ for n fixed at 10 and K fixed at 2	257
Figure 5.45	$J = (\rho^{\wedge}/\sigma[\rho^{\wedge}]) / (\rho^{\wedge}_c/\sigma[\rho^{\wedge}_c])$ under the joint correction for Case 2 [Case A] restriction of range and criterion unreliability as a function of the initial uncorrected effect size estimate and $\rho_{ttY}$ for n fixed at 90 and K fixed at 2	257
Figure 5.46	$J = (\rho^{\wedge}/\sigma[\rho^{\wedge}]) / (\rho^{\wedge}_c/\sigma[\rho^{\wedge}_c])$ under the joint correction for Case 2 [Case A] restriction of range and criterion unreliability as a function of the initial uncorrected effect size estimate and $\rho_{ttY}$ for n fixed at 10 and K fixed at 4	258
Figure 5.47	$J = (\rho^{\wedge}/\sigma[\rho^{\wedge}]) / (\rho^{\wedge}_c/\sigma[\rho^{\wedge}_c])$ under the joint correction for Case 2 [Case A] restriction of range and criterion unreliability as a function of the initial uncorrected effect size estimate and $\rho_{ttY}$ for n fixed at 90 and K fixed at 4	258
Figure 6.1	The estimation of the probability of success conditional on X [i.e. $P[Y^{\wedge} \geq Y_c   X=X_c]$ ]	267

Figure 6.2	The ratio $\Delta = \beta[Y X]/\beta[y x]$ [delta] as a function of $\rho[x,z]$ and $\rho[y,z]$ for $\rho[x,y]$ fixed at 0.10 and $K$ fixed at 2	298
Figure 6.3	The ratio $\Delta = \beta[Y X]/\beta[y x]$ [delta] as a function of $\rho[x,z]$ and $\rho[y,z]$ for $\rho[x,y]$ fixed at 0.30 and $K$ fixed at 2	298
Figure 6.4	The ratio $\Delta = \beta[Y X]/\beta[y x]$ [delta] as a function of $\rho[x,z]$ and $\rho[y,z]$ for $\rho[x,y]$ fixed at 0.60 and $K$ fixed at 2	299
Figure 6.5	The ratio $\Delta = \beta[Y X]/\beta[y x]$ [delta] as a function of $\rho[x,z]$ and $\rho[y,z]$ for $\rho[x,y]$ fixed at 0.80 and $K$ fixed at 2	299
Figure 6.6	The ratio $\Delta = \beta[Y X]/\beta[y x]$ [delta] as a function of $\rho[x,z]$ and $\rho[y,z]$ for $\rho[x,y]$ fixed at 0.10 and $K$ fixed at 4	300
Figure 6.7	The ratio $\Delta = \beta[Y X]/\beta[y x]$ [delta] as a function of $\rho[x,z]$ and $\rho[y,z]$ for $\rho[x,y]$ fixed at 0.30 and $K$ fixed at 4	300
Figure 6.8	The ratio $\Delta = \beta[Y X]/\beta[y x]$ [delta] as a function of $\rho[x,z]$ and $\rho[y,z]$ for $\rho[x,y]$ fixed at 0.60 and $K$ fixed at 4	301
Figure 6.9	The ratio $\Delta = \beta[Y X]/\beta[y x]$ [delta] as a function of $\rho[x,z]$ and $\rho[y,z]$ for $\rho[x,y]$ fixed at 0.80 and $K$ fixed at 4	301
Figure 6.10	The ratio $\Delta = \sigma[Y X]/\sigma[y x]$ [Delta], under Case 3[i] [Case Q[i]] restriction of range as a function of $\rho[x,z]$ and $\rho[y,z]$ for $\rho[x,y]$ fixed at 0.10 and $K$ fixed at 2	305
Figure 6.11	The ratio $\Delta = \sigma[Y X]/\sigma[y x]$ [Delta], under Case 3[i] [Case Q[i]] restriction of range as a function of $\rho[x,z]$ and $\rho[y,z]$ for $\rho[x,y]$ fixed at 0.30 and $K$ fixed at 2	305
Figure 6.12	The ratio $\Delta = \sigma[Y X]/\sigma[y x]$ [Delta], under Case 3[i] [Case Q[i]] restriction of range as a function of $\rho[x,z]$ and $\rho[y,z]$ for $\rho[x,y]$ fixed at 0.60 and $K$ fixed at 2	306
Figure 6.13	The ratio $\Delta = \sigma[Y X]/\sigma[y x]$ [Delta], under Case 3[i] [Case Q[i]] restriction of range as a function of $\rho[x,z]$ and $\rho[y,z]$ for $\rho[x,y]$ fixed at 0.80 and $K$ fixed at 2	306
Figure 6.14	The ratio $\Delta = \sigma[Y X]/\sigma[y x]$ [Delta], under Case 3[i] [Case Q[i]] restriction of range as a function of $\rho[x,z]$ and $\rho[y,z]$ for $\rho[x,y]$ fixed at 0.10 and $K$ fixed at 4	307

Figure 6.15	The ratio $\Delta = \sigma[Y X]/\sigma[y x]$ [Delta], under Case 3[i] [Case Q[i]] restriction of range as a function of $\rho[x,z]$ and $\rho[y,z]$ for $\rho[x,y]$ fixed at 0.30 and K fixed at 4	307
Figure 6.16	The ratio $\Delta = \sigma[Y X]/\sigma[y x]$ [Delta], under Case 3[i] [Case Q[i]] restriction of range as a function of $\rho[x,z]$ and $\rho[y,z]$ for $\rho[x,y]$ fixed at 0.60 and K fixed at 4	308
Figure 6.17	The ratio $\Delta = \sigma[Y X]/\sigma[y x]$ [Delta], under Case 3[i] [Case Q[i]] restriction of range as a function of $\rho[x,z]$ and $\rho[y,z]$ for $\rho[x,y]$ fixed at 0.80 and K fixed at 4	308



## LIST OF TABLES

		PAGE
Table 2.1	Observed scores on k parallel measurements by n testees	44
Table 2.2	Summary of internal-consistency reliability estimators	94
Table 3.1	Eleven validation research designs variations as a function of timing of test and performance measurement and nature of selection decision	109
Table 3.2	Differences in the representativeness of predictive and concurrent validation designs in substituting for the actual selection procedure	111
Table 5.1	Possible outcomes of hypothesis testing and their associated probabilities	216
Table 5.2	A possible nominal scale for scaling G to reflect effect of changes in $Z_b$ on the decision on $H_0$	219
Table 6.1	The effect of Equations 6.33 and 6.34 on the expected criterion performance associated with applicants	302
Table 6.2	The effect of Equations 6.33 and 6.34 on the number of applicants qualifying for selection [ $N_s$ ]	302
Table 6.3	The eight possible conditions affecting the effect of corrections for Case 3[i] [Case C[i]] selection on Z on selection decision making based on the conditional probability of success	309

## CHAPTER 1

## INTRODUCTION, RESEARCH OBJECTIVE AND OVERVIEW OF THE STUDY

The purpose of the introductory chapter is to present a reasoned exposition of the necessity and significance of the envisaged research and to provide a formal statement of the research objective. In essence it is argued that traditional measurement and test theory provide an inadequate conceptual framework from which to assess the practical usefulness of psychometric tests in selection decisions. Selection utility analysis theory in contrast provides a more appropriate framework for analysing, describing and explaining selection decisions in terms of their consequences. Although utility analysis theory clearly shows the deficiencies inherent in defining payoff resulting from a selection procedure solely in terms of the validity coefficient [or translations thereof], it still acknowledges the pivotal role of the validity coefficient in the development and evaluation of selection strategies. Effective selection strategies are possible to the extent that the consequences of a selection decision relevant to the person making the selection decision are systematically related to the information used to make the decision. Relevant information therefore constitutes a necessary, although not sufficient, prerequisite to develop an effective, efficient and defensible selection strategy. The fundamental notion of a linear relationship between a predictor [or a linear composite of predictors] and a criterion expressed in terms of the parameters [i.e., regression parameters and correlation coefficient] of the general linear model [GLM] thus forms a basic building block in all utility analysis models. The implicit requirement made by the GLM applied to selection problems is that the context in which the selection strategy will eventually be applied should be mirrored in the context in which the selection strategy is developed and evaluated/trial-tested. This would firstly require the dependent variable being regressed on the weighted composite of predictors to be the latent criterion and not an operational indicator of it attenuated by measurement error. A further requirement is that the parameters of the GLM are estimated from a representative sample from the actual applicant population [i.e. the range of scores observed in the sample should correspond to the range of scores for which the predictor will eventually be used]. The observed criterion variable does, however, contain measurement error and a representative sample from the applicant population is normally not available. The attenuating effect of both these influences on the actual, empirically derived, validity coefficient is well recognised, and procedures to derive a correlation corrected for attenuation due to the unreliability of the criterion and/or restriction of range are generally available [Ghiselli, Campbell & Zedeck, 1981; Lord & Novick, 1968]. If these corrections for attenuation due to the unreliability of the criterion and/or restriction of range would, however, be restricted to the validity coefficient only, relatively little would be gained in as far as the decision function, the different definitions of payoff permitted by utility analysis theory and the statistical audit of the fairness of the selection strategy would remain unaffected. The actual practical usage of the selection procedure, as well as any evaluative conclusion on the usefulness of the selection procedure, would therefore not be affected. This, however, raises the question whether

corrections for attenuation due to the unreliability of the criterion and/or restriction of range should be extended to the GLM based decision function, the different definitions of payoff permitted by the utility analysis theory or the statistical audit of the fairness of the selection strategy, and if so, what the effect of these corrections would be. The question regarding the advisability of correcting for the unreliability of the criterion and/or restriction of range in essence hinges on the probability that such corrections would change decisions, the consequences of these changed decisions and the cost of obtaining the reliability estimates needed to implement the corrections.

## 1.1 INTRODUCTION

Organisations do not constitute natural phenomena but rather man made phenomena and therefore exist, in as far as human behaviour is motivated, goal directed behaviour, for a definite reason and with a specific purpose. The human resource function justifies its inclusion in the family of organisational functions through its commitment to contribute towards organisational goals. Almost all definitions of human resource management formally and explicitly profess this commitment. At the same time, however, insufficient effort is devoted to explicitly delineate these organisational goals in abstract terms with sufficient clarity. Although the general motivation underlying interventions launched under the banner of the human resource function [e.g. human resource selection] is thus clear, inadequate explication of the primary organisational goal still has the effect of obscuring the logic in terms of which the human resource function should be able to justify its assertion that it contributes towards organisational goals. Consequently it becomes very difficult to infer, through a reasoned argument, the appropriate criteria in terms of which human resource intervention [e.g. human resource selection] should be evaluated.

In order to be instrumental in the satisfaction of the multitude of needs of society, organisations have to combine and transform scarce factors of production into products and services with maximum economic utility. The organisation is thereby confronted with a choice of alternative utilisation possibilities regarding the limited factors of production it has access to. The organisation is guided in this choice by the economic principle, which commands, on behalf of society, the organisation to attain with the lowest possible input of production factors the highest possible output of need satisfying products and/or services. The organisation [in a capitalistic system] complies with the demand of the economic principle because such compliance enables it to maximise its profits. The motivation for the organisation to serve society through the efficient production of need satisfying products and/or services thus lies in the opportunity to utilise the capital it has to its disposal, via economic activities directed at the creation of need satisfying products and/or services, for its own benefit. In order to have an optimal exploitation of this opportunity, however, profitability maximisation must be

designated as the primary organisational goal. The primary objective for the organisation thus is the maximisation of the profit earned over a particular period relative to the capital used to generate that profit. Specifically this objective refers to the rate of return on equity [Rademeyer, 1983], that is profit [expressed as a percentage] earned on capital of ordinary shareholders. By subscribing to this principle the organisation thus commits itself to the maximisation of the value of the organisation for its owners [shareholders] as it manifests itself in the price of its share on the stock exchange [in the case of listed companies] and the magnitude of dividends declared [Clark, Hindelang & Prichard, 1984; Levy & Sarnat, 1994; Lumby, 1994; Rademeyer, 1983]. Financial decisions, especially decisions regarding the appropriation of capital, are guided by the quest to add value to the organisation. Therefore, in order to justify the investment of capital in any project in terms of the quest to add value to the organisation, the expected return on the investment, or the expected cash inflow generated by the project, should exceed the amount initially invested in the project. In estimating the rationality of a contemplated investment the time value of money [the fact that cash inflows occur over the lifetime of the project] must be taken into account as well as the risk associated with the investment [Levy & Sarnat, 1994].

In order to actualise the primary objective of the organisation a multitude of mutually coordinated activities need to be performed which can be categorised as a system of inter-related organisational functions. The human resource function represents one of these organisational functions. The human resource function aspires to contribute towards organisational objectives through the acquisition and maintenance of a competent and motivated work force, as well as the effective and efficient utilisation of such a work force [Crous, 1986]. The importance of human resource management flows from the basic premise that organisational success is significantly dependent on the quality of its work force and the way the work force is utilised and managed. Labour constitutes a pivotal production factor due to the fact that the organisation is managed, operated and run by people. Labour is the life giving production factor through which the other factors of production are mobilised and thus represents the factor which determines the effectiveness and efficiency with which the other factors of production are utilised [Marx, 1983]. The management of human resources is, however, complicated by the intricate, and to a certain extent enigmatic, nature of working man as the carrier of labour as production factor. This leads to the basic premise that credible and valid theoretical explanations [i.e. social science theory] for the different facets of the work behaviour of working man constitute a fundamental and indispensable, though not sufficient, prerequisite for efficient and equitable human resource management.

Industrial/Organisational Psychology embodies the conviction that, in spite of the extreme complexity of human behaviour, regularities underlying the work-related behaviour of working man can be unraveled and explained in terms of a nomological network of constructs [i.e. theory]. According to Veldsman [1986] Industrial/Organisational Psychology is the behavioral science directed at the psychological explanation of the work-related behaviour of working man. Industrial/Organisational



Psychology derives the fundamental reason for its existence from the potential "therapeutic" value [Mouton & Marais, 1985] of valid social science theory. Industrial/Organisational Psychology would concede that perfect understanding and complete certainty regarding the principles governing the work behaviour of man represents an unattainable ideal. Industrial/Organisational Psychology would, however, still contend that sufficiently comprehensive approximations of reality can be achieved through a scientific methodology [therein lies in part the relevance of I/O Psychology's commitment to the scientific method of inquiry] to be of significant relevance to practical human resource decisions [Milkovich & Boudreau, 1994; Mouton & Marais, 1985].

To the extent that Industrial/Organisational Psychology can produce credible and valid theoretical explanations for the different facets of the work behaviour of working man, an opportunity exists to derive, through deductive inference, practical human resource interventions designed to affect either employee flows or employee stocks [Boudreau, 1991; Milkovich & Boudreau, 1994]. Interventions designed to affect employee flows attempt to change the composition of the work force by adding, removing or reassigning employees [e.g. through recruitment, selection, turnover, or internal staffing] with the expectation that such changes will manifest in improvements in work performance. In contrast interventions designed to affect employee stock attempt to change the characteristics of the existing work force in their current positions or the work situation itself [e.g. through training, performance feedback, compensation or job redesign]. The expectation is that such changes will manifest in improvements in work performance [Boudreau, 1991]. Improvements in work performance are affected through increases in work force quality which in turn are brought about by the aforementioned two types of human resource interventions. Improvements in work performance as such would however not constitute sufficient evidence to justify the intervention that affected these improvements. Given that the human resource function's inclusion in the family of organisational functions is justified through its commitment to contribute towards the primary organisational objective of maximising the value of the organisation for its owners, it logically follows that all interventions initiated by the human resource function should, in the final analysis, also be evaluated with the yardstick of profitability. Babbel, Stricker and Vanderhoof [1994, p. 3] come to a similar conclusion with regards to insurance managers:

One of the most basic tenets of modern financial theory is that managers should act in a manner consistent with maximizing the value of owners' equity. While there are theoretical conditions under which this tenet may not always apply, for practical purposes companies usually espouse it as a financial goal. If an insurer accepts this maxim as a company goal, it follows that the firm should view the performance of insurance managers and operatives in terms of whether this performance helps to promote higher firm value.

The design, implementation and operation of human resource interventions thus only make sense from an institutional perspective if a satisfactory [appropriately discounted] return on the capital invested in

the intervention is achieved over the period in which the intervention generates its effect. There thus rests an obligation on the human resource function to prove through appropriate financial indicators [Boudreau, 1991; Cronshaw en Alexander, 1985] that its interventions do add value to the organisation [Cascio, 1991b]. Furthermore it seems reasonable to contend that the only rational way the human resource function can compete for limited capital on a more or less equal footing with the other organisational functions is in terms of expected returns on capital invested [Cronshaw en Alexander, 1985]. The burden of persuasion rest particularly heavy on the human resource function due to its general inability in the past to demonstrate its ability to contribute to bottom-line success [Cascio, 1991b]. In as far as the human resource function had neglected to meet this burden of persuasion, a relative lack in stature, influence and recognition, in comparison to the other organisational functions, seems hardly surprising [Cronshaw en Alexander, 1985; Gow, 1985; Sheppeck & Cohen, 1985]. Fitz-Enz [1980, p. 41] comments as follows in this regard:

Few human resource managers, even the most energetic, take the time to analyze the return on the corporation's personnel dollar. We feel we aren't valued in our own organizations, that we can't get the resources we need. We complain that management won't buy our proposals and wonder why our advice is so often ignored until the crisis stage. But the human resources manager seldom stands back to look at the total business and ask: Why am I at the bottom looking up?. The answer is painfully apparent. We don't act like business managers, like entrepreneurs whose business happens to be people.

The question regarding the return on capital invested in human resource interventions should, however, not be exclusively addressed to the human resource practitioner but should also be directed to the industrial psychologist as behavioral scientist. Industrial/Organisational Psychology studies the behaviour of working man in an effort to try and uncover the principles governing work related behaviour on account of the "therapeutic" value such insight offers. To the extent that it can be shown that human resource interventions, deductively derived from Industrial/Organisational theory, do [or do not] add value to the organisation, crucial information is fed back to the [basic/academic] research arena. Apart from the guidance value thereof, such feedback, in the final analysis, constitutes the decisive criterion in terms of which Industrial/Organisational Psychology [like its human resource management counterpart] should judge the extent to which it succeeds in its professed mission.

Human resource interventions can, however, not be evaluated solely in terms of the return on capital invested in the intervention, since such interventions not only impact on the primary organisational objective of maximising the value of the organisation for its owners. Human resource interventions also impact on the psychological, physical and social wellbeing of current and prospective employees. Human resource interventions not only have an institutional payoff, but also an individual payoff. Equal access to human resource intervention opportunities for all current and aspirant employees would, from an institutional perspective, be considered irrational since it would nullify any institutional

payoff that could otherwise have been derived from such interventions. Individual employees will, therefore, unavoidably derive differential benefit from human resource interventions. Because of the disparate impact of human resource interventions it becomes imperative that the fairness or justice of such actions be assessed so as to ensure equitable human resource practices [Singer, 1993].

Human resource interventions therefore impact either positively or negatively on the financial position of the organisation and on the lives of the people that serve as its employees or aspire to do so. Two basic criteria are thus implied in terms of which human resource interventions need to be evaluated, namely efficiency and equity [Milkovich & Boudreau, 1994]. Two stakeholders are furthermore thereby implied, namely management, representing the [financial] interests of the owners, and labour and its organised representatives. Should the human resource function be challenged by management and/or by organised labour to defend any intervention it would, therefore, have to lead evidence to show the efficiency and equity of the disputed intervention. The extent to which the human resource function would be able to establish the integrity of the challenged intervention would depend on the validity and credibility of the evidence led in its defense. The validity and credibility of the verdict on the efficiency and equity of an intervention in turn would largely depend on the methodology in terms of which the evidence was generated. Therein lies part of the motivation for the commitment of Industrial/Organisational Psychology to the scientific method of inquiry.

## 1.2 PERSONNEL SELECTION

Selection, as it is traditionally interpreted, represents a critical human resource intervention in any organisation in as far as it regulates the movement of employees into, through and out of the organisation. As such selection thus firstly represents a potentially powerful instrument through which the human resource function could add value to the organisation [Boudreau, 1983b; Boudreau & Berger, 1985a; Cascio, 1991b; Cronshaw en Alexander, 1985]. Selection secondly, however, also represents a relatively visible mechanism through which access to employment opportunities are regulated. Because of this latter aspect selection, more than any other human resource intervention, had been singled out for intense scrutiny from the perspective of fairness and affirmative action [Arvey & Faley, 1988; Milkovich & Boudreau, 1994]. The aforementioned two basic criteria [efficiency and equity] thus also apply to the evaluation of selection procedures [Milkovich & Boudreau, 1994]. The quest for efficient and equitable selection procedures, however, if not left to chance but pursued in a rational and judicious manner, necessitates a clear and comprehensive elucidation of all relevant factors that affect these desired outcomes. Only if guided by a formal framework synthesising all relevant selection parameters and the way these parameters interact, can diagnostic analyses/psychometric



audits of selection procedures, motivated by the desire to improve the efficiency and equity of such procedures, become a viable and worthwhile endeavor.

Decision theory seems to provide the most productive framework in which to cast the selection problem so as to attain such a guidance system [Boudreau, 1991; Cascio, 1991b; Cronbach & Gleser, 1965; Huber, 1980]. Cronbach and Gleser [1965], in their seminal and pioneering publication, vigorously advocated the inability of traditional measurement and test theory, due to its emphasis of the instrument and precision of measurement, to provide an adequate conceptual framework from which to assess the practical usefulness of tests in selection. In selection, or any other applied situation for that matter, psychological tests essentially serve the purpose of providing information for decision making. The focus thus should be on the quality of the decision making and not on the psychometric properties of the test, although this should not be interpreted to mean that measurement and test theory in the Gulliksen [1950] tradition are to be regarded as irrelevant and obsolete. Two general decision-theoretic approaches towards selection can be differentiated [Schuler & Guklin, 1991]:

- a descriptive approach focusing on explicating the decision heuristics actually used by decision makers; and
- a normative approach focusing on explicating a formal decision structure to be used by decision makers as a decision aid [Huber, 1980] to compensate for their bounded rationality [March & Simon, 1958] when considering alternative options.

Selection can be conceptualised in a normative decision-theoretic framework from either an operational or strategic perspective. Thus the focus could be placed on the individuals about whom decision are required or on the selection procedure itself.

Cascio [1991, p. 178], in an effort to establish the capacity of decision theory to provide a suitable framework in which to cast the selection problem, clearly approaches selection from an operational perspective when he asserts:

It should serve as some comfort to know that all personnel decisions can be characterized identically. In the first place there is an individual about whom a decision is required. Based on certain information about the individual [for example performance appraisals, assessment center ratings, a disciplinary report], decisions makers may elect to pursue various alternative courses of action.

Cronbach and Gleser [1965, pp. 135-136], likewise, in their endeavor to point out the inadequacy of traditional measurement theory as a conceptual vehicle to evaluate the usefulness of selection instruments, operate from an operational perspective when they state:



The traditional theory views the test as measuring instrument intended to assign accurate numerical values to some quantitative attribute of the individual. It therefore stresses, as the prime value, precision of measurement and estimation. .... In pure science it is reasonable to regard the value of a measurement as proportional to its ability to reduce uncertainty about the true value of some quantity. .... In practical testing, however, a quantitative estimate is not the real desideratum. A choice between two or more discrete treatments must be made. The tester is to allocate each person to the proper category, and accuracy of measurement is valuable only insofar as it aids in this qualitative decision.

### 1.2.1 An Operational Perspective on Personnel Selection

When viewed from an operational perspective selection decisions can be characterised in terms of the following structural elements [Cronbach & Gleser, 1965]:

- a set of individuals about whom limited information is available
- a set of decision options or treatments to which individuals have to assigned;
- a decision function specifying treatment assignment contingent on information about individuals;
- a set of outcomes contingent on the assignment of individuals to specific treatments described in terms of a multidimensional [composite] criterion; and
- a utility scale on which the possible outcomes are evaluated.

Assuming that a selection quota is in force, the task with which the selection decision maker is confronted is in essence to identify a subgroup from the total group of applicants to allocate to the accept treatment, based on limited but relevant information about the applicants. The subgroup, furthermore, has to be chosen so as to maximise the average gain on the utility scale on which the outcomes of decisions are evaluated. The utility scale/payoff and the actual outcomes or ultimate criterion [Austin & Villanova, 1992] are the focus of interest in selection decisions [Ghiselli, Campbell & Zedeck, 1981]. These are, however, by definition not available at the time of the selection decision. Under these circumstances, and in the absence of any [relevant] information on the applicants, no possibility exists to enhance the quality of the decision making over that that could have been obtained by chance. The only alternative to random decision making [other than not to take any decision at all], would be to predict expected outcomes/criterion performance [or expected utility] actuarially [or clinically] from relevant, though limited, information available at the time of the selection decision. Thus some substitute<sup>1</sup> for the criterion is called for [Ghiselli, Campbell & Zedeck, 1981]. The only information available at the time of the [fixed treatment] selection decision [Cronbach & Gleser, 1965]

---

<sup>1</sup> Formally  $X$ , and therefore by implication  $E[Y|X]$ , could be considered a substitute for  $Y$  if and to the extent that  $|\rho[X,Y]| > 0$  [ $p < 0.05$ ] and if measures of  $X$  can be obtained at the time of or prior to the selection decision

that could serve as such a substitute would be psychological, physical, demographic or behavioural information on the applicants. Such substitute information would be considered relevant to the extent that the regression of the [composite] criterion [or corresponding utility scale values] on a weighted [probably, but not necessarily, linear] combination of information explains variance in the criterion. Thus the existence of a relationship, preferably one that could be articulated in statistical terms, between the outcomes considered relevant by the decision maker and the information actually used by the decision maker, constitutes a fundamental and necessary, but not sufficient, prerequisite for effective and equitable selection decisions.

Two, and only two, options exist in terms of which an acceptable substitute for the criterion could be found. In terms of the first option the job in question would be systematically analysed via one or more of the available job analysis techniques [Gatewood & Feild, 1994] to identify and define the behaviours that collectively denote job success if exhibited on the job. Substitute information would then be obtained through low or high fidelity simulations of the job content. These simulations in a selection context necessarily occur off the job and at prior to the selection decision. Such simulations would elicit behaviour that, if it would in future be exhibited on the job, it would denote a specific level of job performance. These behaviours are sometimes referred to as competencies [Saville & Holdsworth Ltd, 1996]. If competencies, defined in this sense, are assessed off the job via some form of simulation [in contrast to on the job via actual job performance] the resultant assessments combined can be regarded as a predictor of the criterion and thus as a substitute for the criterion. If, however, competencies, defined in the above interpretation of the term, would be assessed on the job via actual job performance the resultant assessments combined would constitute criterion measures.

In terms of the second option the job in question would also be systematically analysed but now with the purpose of inferring suppositious critical incumbent attributes, believed to be determinants of the level of criterion performance that would be attained, from the description of the job content and context. These critical attributes are unfortunately also sometimes referred to as competencies [Spangenberg, 1990] thereby creating a good measure of confusion, misunderstanding and discord in contemporary psychometric debate. The presumed interrelationship between these hypothesised determinants and the way they collectively combine in the criterion is postulated in a nomological network or latent structure [Campbell, 1991; Kerlinger, 1986] as a complex hypothesis explaining criterion performance in the job in question. These hypothesised determinants of criterion performance, or a person centered subset thereof, could, to the extent that the tentative performance theory is indeed valid, serve in combined form as a suitable substitute measure for the, still to be realised, actual criterion scores. The way these hypothesised determinants of performance should be combined is suggested by the way these determinants are linked in the postulated nomological network.

The extent to which effective substitute criterion measures are obtained through these two options should be the subject of empirical validation investigations. This would establish the foundation for a comprehensive argument in terms of which the actual use of any selection procedure could be justified from the perspective of both efficiency and equity.

In the case of the first option, the fundamental question is whether a content [factorially] valid description of the competencies are obtained through the simulation and, if so, whether this description correlates statistically significant with an independently obtained construct valid measure of the ultimate criterion [Austin & Villanova, 1992].

In the case of the second option, the fundamental question is whether a weighted linear [or even possibly non-linear, depending on the directive emanating from the postulated nomological network] combination of construct valid measures of the hypothesised determinants of performance significantly explains variance in a construct valid measure of the ultimate criterion.

Clearly substantial differences exist between the logic underlying these two options in terms of which substitute criterion measures are generated. Most relevant, is the fact that the second option necessitates the explication of an underlying performance theory whilst the first option can proceed without any significant understanding as to why inter-individual performance differences exist. Both arguments, however, maintain that effective, though not necessarily efficient, selection is contingent on the identification of a substitute [in the form of a differentially weighted combination] for the ultimate criterion which shows a statistically describable relationship with an operational measure of the ultimate criterion. Both arguments, furthermore, contend that the same condition constitutes a necessary, though not sufficient condition to achieve fair or equitable employee selection. The extent to which the substitute succeeds in representing the ultimate criterion is in both cases described by the validity coefficient as a multiple correlation between a composite intermediate criterion and a weighted combination of the indicators of performance [ $R_{yy'}$ ]. A perfect correlation would imply that the decision maker has a flawless understanding of the predictor-criterion latent structure [Campbell, 1991], can obtain psychometrically flawless measures of all relevant constructs and thus could, with perfect precision and complete certainty, infer values on the intermediate criterion from the combined substitute measure. The decision maker would have a relative simple selection problem to contend with if this would be the case because it would imply that he can, with complete certainty, anticipate the actual outcomes for any applicant should such an applicant be accepted. For such a perfectly informed decision maker there would be no unanticipated consequences and thus also no risk and no decision errors. This is, however, never the case [March & Simon, 1958]. The situation the decision-maker has to contend with is characterised by relevant, limited and psychometrically flawed information. Relevant, limited and psychometrically flawed information would thus imply the classic imperfectly correlated, bivariate [linear, homoscedastic and normal] distribution of composite criterion and composite



predictor/substitute criterion scores [Boudreau, 1991; Campbell, 1991; Cronbach & Gleser, 1965]. This classic validity model forms the foundation of all selection procedures [Boudreau, 1991; Campbell, 1991]. The decision maker's lack of perfect understanding and complete certainty regarding the principles governing the outcomes resulting from acceptance, combined with his reliance on fallible information, thus denies him the possibility of anticipating selection outcomes with complete certainty. His access to relevant but limited and psychometrically flawed information, however, still permits him the possibility of statistically describing the conditional distribution of outcomes. Thus the decision maker can only base his decision whether to accept an applicant on the expected outcome conditional on information on the applicant or, if a minimally acceptable outcome state can be defined, the conditional probability of success [or failure] given information on the applicant. Alternatively, the bivariate distribution could be converted into a contingency table through the formation of intervals on both the predictor and the criterion. The resultant validity matrix [Cronbach & Gleser, 1965] or expectancy table [Ghiseli, Campbell & Zedeck, 1981; Lawsche & Balma, 1966], indicating the probability of a specific criterion state conditional on a specific information category, could then be used as basis for decision-making. If one would, furthermore, be willing to assume the existence of a cardinal utility scale on which outcomes can be evaluated, if not linearly, at least monotonically related to the criterion scale, it becomes possible to translate the expected outcome to the expected payoff conditional on information on the applicant. [Crocker & Algina, 1986; Cronbach & Gleser, 1965].

The embryonic question, from which a construct orientated [option-two] selection procedure is ultimately conceived, asks with deceptive simplicity why differences in performance exist. Inability to answer this question in terms of a valid performance theory effectively eliminates the possibility to differentiate between better and poorer employment prospects. Unless one is willing to assess suitability for employment through demonstration of competency in tasks corresponding to the critical performance areas comprising the job in question [i.e. through simulations via a content orientated approach] and thus can sidestep the question as to why performances differences exist. Attaching the term ability to each critical performance area provides no satisfactory answer to this question but simply creates the impression of explanation. This line of reasoning suggests the validation of a selection procedure to be, in essence, a form of applied explanatory scientific research [Binning & Barrett, 1989; Ellis & Blustein, 1991; Landy, 1986]. Landy [1986, pp. 1187-1188] supports this assertion, by stating:

The validity analyst is carrying out traditional hypothesis testing. At least by implication, the hypothesis being considered is of the following form: People who do well on test X will do well on activity Y, or  $Y=f[X]$ . Investigators should not lose sight of the fact that validity studies are attempts to develop a theory of performance that explains how an individual can [or will] meet the demands of a particular job.

This seemingly benign position, that the validation of a selection procedure should be viewed from a scientific, theory building and hypothesis testing perspective, has in fact a multitude of far reaching implications. By adopting this position one, in essence, formally acknowledges three basic premises, two of which were already alluded to in the preceding discussion, namely:

- valid theoretical explanations for the different facets of the work behaviour of working man constitute a fundamental and indispensable, though not sufficient, prerequisite for effective and efficient human resource management interventions;
- the validity of theoretical explanations for the different facets of the work behaviour of working man depend on the quality and thoroughness of the theorising that, in reaction to the research initiating [why] questions, bring forth the theoretical explanations; and
- the credibility/validity of statements on the validity of theoretical explanations for the different facets of the work behaviour of working man depend on the unassailability of the methodological argument in terms of which these statements are justified.

Thus, by adopting this research orientated stance, one unequivocally restores theorisation and conceptualisation to their rightful place in the selection procedure development process [Binning & Barrett, 1989; Campbell, 1991; McGuire, 1997]. More crucial, however, for the argument in question, is the formal recognition that the credibility/validity of statements on the validity of explanations of the basic form  $Y=f[X]$  depends on the scientific rationality of the methodological argument through which such conclusions were arrived at. This in turn, clearly has important implications for the type of evidence the human resource function should lead should it be summoned, either by management or regulatory agencies, to defend its selection procedures in terms of its efficiency and equity respectively. The ability of the human resource function to successfully defend its selection procedure, however, is not only dependent on the credibility/validity of statements on the validity of the performance hypothesis underlying the selection procedure. Apart from the fact that further utility and fairness analyses are required to establish the efficiency and equity of a selection procedure, a more fundamental issue to be considered first, is the extent to which the validation research evidence can be logically transferred to the actual area of application. Both these issues naturally emerge as critical themes if the concept of research design is formally introduced.

Research design, if more extensively defined than is usually the case [Kerlinger, 1986], refers to the plan and structure of the entire investigative process [Kerlinger 1986; Mouton & Marais, 1985] so conceived as to justify the actual application of the selection procedure. Internal and external validity constitute two general criteria of research design [Campbell & Stanley, 1963; Kerlinger, 1986]. Due to the more extensive meaning attached to the research design concept, a concomitantly more comprehensive definition of internal validity, encompassing the conventional interpretation of the term [Campbell & Stanley, 1963; Cook & Campbell, 1979], is also required. Internal validity is typically interpreted to mean the confidence with which [between group or systematic] variance in the dependent variable of

interest can be attributed to the effect of one or more independent variables [Babbie, 1989; Cook & Campbell, 1979]. A more encompassing term is, however, needed here to refer to the confidence with which the latent structure postulated by the research hypothesis [and not the operational hypothesis] may be regarded as corroborated by the research evidence marshaled in its support within the limits set by the specific conditions prevailing at the time of the investigation. However, rather than willy-nilly [and somewhat arrogantly] redefining Campbell and Stanley's [1963] concept of internal validity to satisfy the current argument's need for a single, suitably encompassing, concept, the trilogy of statistical design criteria introduced by Cook, Campbell and Peracchio [1991] are employed to represent the pertinent considerations. Statistical conclusion validity, internal validity and construct validity [Cook, Campbell & Peracchio, 1991] collectively determine the credibility/validity of statements on the validity of the performance hypothesis underlying the selection procedure. The confidence with which the latent structure postulated by the research hypothesis may be regarded as corroborated by the research evidence marshaled by a validation study would, therefore, be threatened by;

- the lack of inferential validity of the deductive argument in terms of which the research hypothesis is operationalised, especially the construct [and/or content] validity of the measurement procedures referred to in the premises contained in the argument [Mouton & Marais, 1985];
- the extent to which the research design fails to controls variance, that is, the extent to which the design fails to maximise experimental/systematic variance, fails to minimise error variance and fails to control extraneous variance [Campbell & Stanley, 1963; Kerlinger, 1986]; and
- the extent to which statistical power [Cohen 1977] is decreased.

If validation of a selection procedure is viewed from a scientific, theory-building and hypothesis testing perspective, the preceding argument makes it painfully clear that the "stamp collecting" approach [Landy, 1986, p. 1184], induced by the tripartite division of validity [APA, AERA & NCME, 1985], represents an unjustifiable and misleading practice. To paraphrase Guion [1980, p. 386], there exists only one road, and not "three different roads, to psychometric salvation" within the context of selection research. The validation of multiple inferences are required to scientifically justify the pivotal tenet on which selection procedures are based [Binning & Barrett, 1989]. Landy [1986, pp. 1185-1188] summarises the unificationist perspective on the validation of selection procedures as follows:

The labels *content*, *construct*, and *criterion-related* are not completely useless, nor are they interchangeable. They had their use in 1954, and they have their value in 1986. However, their value is not as types of validity. Instead, their value is in pointing out that there is more than one type of inference that can be made from a test score. .... Aspects of validity cannot be easily separated from one another. Because the words *content*, *criterion-related*, and *construct* can be used as aids in discussion, one should not be seduced into thinking of those



words as standing for discrete and independent processes. Instead, the words simply represent parts of a larger system that addresses the goal of hypothesis testing.

A validation design [interpreted in the wider sense of the term] with adequate statistical conclusion validity, internal validity and construct validity constitutes a necessary, but not sufficient, requirement to construct a rigorous argument, that would be able to successfully withstand intellectual onslaught, should it be required in litigation. A second, very crucial, line of defense required to defend the core of any selection procedure concerns the external validity of the validation design.

To a certain extent external validity represents a rather elusive concept to pin down with a satisfactory constitutive definition. For the purpose of the present argument, external validity is interpreted to refer to the degree of confidence with which the [internally valid] results of a specific empirical validation study can be generalised [or transported] to a specific area of application [Cook & Campbell, 1979; Stanley & Campbell, 1963]. Generalisation, furthermore, always represents a problematic endeavor, because it entails a risky extrapolation beyond the specifics characterising the validation study [Cook & Campbell, 1979]. Generalisation, or induction, can never be completely justified by the evidence supplied by a validation study; strictly speaking generalisation beyond the limits set by the validation study is not possible at all [Mouton & Marais, 1985; Stanley & Campbell, 1963]. Nevertheless, given the applied nature of selection validation research, an attempt at generalisation is unavoidable. According to Stanley & Campbell [1963] external validity is threatened by the potential specificity of the demonstrated effect of the independent variable[s] to particular features of the research design not shared by the area of application. In selection validation research the effect of the [composite] independent variable on the criterion is captured by the validity coefficient. The area of application is characterised by a sample of actual applicants drawn from the applicant population and measured on a battery of fallible predictors with the aim of "estimating their actual contribution to the organisation [i.e. ultimate criterion scores] and not an indicator of it attenuated by measurement error" [Campbell, 1991, p. 694] from a weighted linear composite of predictors derived from a representative sample from the actual applicant population. The question regarding external validity, in the context of selection validation research, essentially represents an inquiry into the unbiasedness of the parametric validity coefficient estimated from the sample statistic obtained through the validation study. The parameter of interest is the correlation coefficient obtained when the sample weights derived from a representative sample are applied to the applicant population and the weighted composite score is correlated with the criterion, unattenuated by measurement error, in the population [Campbell, 1991].

The preceding discussion clearly indicates the term "applicant population" to be of central importance should a sufficiently precise depiction of the area of actual application be desired. The term "applicant population", however, even if defined as the population to which a selection procedure will be applied, still has an annoying impreciseness to it. A more unambiguous definition of the term, however,

depends on how the selection procedure would be positioned relative to any selection requirements already in use [i.e. whether it would replace, follow on, or be integrated with current selection requirements]. This issue, moreover, is linked to the question regarding the appropriate decision alternative with which to compare the envisaged selection procedure when examining its strategic merit. Further clarification of the term will consequently be attempted in the ensuing discussion on the strategic evaluation of selection procedures. In the context of selection validation research, given the aforementioned depiction of the area of application, the following specific threats to external validity could thus be identified [Campbell, 1991; Lord & Novick, 1968; Tabachnick & Fidell, 1989]:

- the extent to which the actual or operationalised criterion contains random measurement error;
- the extent to which the actual or operationalised criterion is systematically biased; i.e. the extent to which the actual criterion is deficient and/or contaminated [Blum & Naylor, 1968];
- the extent to which the validation sample is an unrepresentative, biased, sample from the applicant population in terms of homogeneity and specific attributes [e.g. motivation, knowledge/experience];
- the extent to which the sample size and the sample size to number of predictors ratio allow capitalisation on chance and thus overfitting of data.

The conditions listed as threats do all affect the validity coefficient [Campbell, 1991; Crocker & Algina, 1986; Dobson, 1988; Hakstian, Schroeder & Rogers, 1988; Lord & Novick, 1968; Mendoza & Mumford, 1987; Messick, 1989; Olsen & Becker, 1983; Schepers, 1996]. Some consistently exert upward pressure, others downward pressure and for some the direction of influence varies. It thus follows that, to the extent that the aforementioned threats did operate in the validation study but do not apply to the actual area of application, the obtained validity coefficient cannot, without formal consideration of these threats, be generalised to the actual area of application. The obtained validity coefficient thus cannot, without appropriate corrections, be considered an unbiased estimate of the actual validity coefficient of interest. Campbell [1991, p. 701] consequently recommends that:

If the point of central interest is the validity of a specific selection procedure for predicting performance over a relatively long time period for the population of job applicants to follow, then it is necessary to correct for restriction of range, criterion unreliability, and the fitting of error by differential predictor weights. Not to do so is to introduce considerable bias into the estimation process.

Information is obtained and analysed/processed at a cost with the purpose of making a decision [i.e. choosing between two or more alternatives] which results in outcomes with certain value to the decision-maker. Additional information and/or additional analyses of information could be considered rational if it results in an increase in the value of the outcomes at a cost lower than the increase in value.



The foregoing argument thus implies that corrections applied to the obtained correlation coefficient are rational to the extent that [Boudreau, 1991]:

- the corrections change decisions on the validity of the research hypothesis in that they change the magnitude of either the validity coefficient, or the standard error of the validity coefficient, or both [or at least the a priori probability of rejecting  $H_0$  assuming  $H_0$  to be false];
- the change in decisions have significant consequences;
- the cost of applying the statistical corrections are low.

If the probability of these corrections changing the pre-correction conclusion on the validity of the research hypothesis is small, and the value of the change in decision is less than the cost of implementing the corrections, one could thus contend that very little justification actually exists for these corrections.

The concept of external validity as used here should, however, not be confused with two closely affiliated, but distinct concepts, namely generalisability theory [Cronbach, Gleser, Nanda & Rajaratnam, 1972] and validity generalisation theory [Schmidt & Hunter, 1977; Schmidt, Hunter, Pearlman & Hirsh, 1985]. Generalisability theory, as an extension of reliability theory, is a statistical theory on the accuracy of generalisations from a testee's observed score on some measure to the average score that person would have received over all acceptable observations [Shavelson & Webb, 1991]. Validity generalisation theory postulates, in contrast to the traditional situational specificity position, that observed situational variability in test-criterion correlations can be explained in terms of statistical artifacts and thus that criterion-related validity is generalisable across different situations [Messick, 1989]. The statistical artifacts normally considered by validity generalisation theory to contribute to artifactual validity variance are sampling error, criterion unreliability, predictor unreliability, restriction of range and criterion contamination [Duvenage, 1991; Messick, 1989;]. Validity generalisation theory clearly also represents a manifestation of the Cook and Campbell [1979] concept of external validity. It differs, however, from the argument developed here, in that it requires meta-analysis [Hunter, Schmidt & Jackson, 1982] of extensive criterion-related validity data sets, generated by individual validation studies, to estimate the mean and standard deviation of the true validity distribution. The present argument, on the other hand, focuses on the removal of artifactual effects in the individual validation study result. Neither an extreme validity generalisability position nor an extreme situational specificity position is, however, probably warranted. The two positions should thus not be pitted against each other, but should rather be viewed as interdependent, mutually complementary, approaches [Messick, 1989; Raudenbush & Bryk, 1985]. Although an extreme validation generalisability position would render local validation studies questionable, the extensive benefits afforded by such studies, point to the necessity and pertinence of local validation studies in conjunction with validity generalisation research [Messick, 1989; Raudenbush & Bryk, 1985; Schmidt, Hunter, Pearlman & Hirsh, 1985].

Once the latent structure, postulated by the research hypothesis as foundation for the selection procedure, is adequately corroborated by the research evidence marshaled in its support and the results can, following appropriate corrections, be generalised [or transported] to the specific area of application, attention shifts to the selection strategy.

A selection strategy or selection decision function refers to an explicit or implicit rule dictating, conditional on available information, whether an applicant should be terminally rejected or accepted, or whether and what further information should be collected [Cronbach & Gleser, 1965; Gatewood & Feild, 1994; Muchinsky, 1983]. A selection strategy, in essence, consists of conditional probabilities [normally 0 or 1, but not necessarily] stating the probability of each treatment/decision, conditional on information on the applicant. A selection strategy thus could be enunciated in terms of a strategy matrix [Cronbach & Gleser, 1965], prescribing, or describing, through conditional probabilities associated with different information categories, how applicants should be, or are, assigned to the different possible treatments.

The nature of the selection strategy depends on the manner in which multiple selection information is combined for decision-making [that is, whether compensation will be permitted, whether multi-stage sequential testing will be employed] and whether quotas are in force. Whether compensation is permitted would impact on the number of ways [Coxon, 1982; Schiffman, Reynolds & Young, 1981] comprising the strategy matrix, restricting it to two if compensation is allowed. Assuming an information x decision format, multi-stage sequential testing would increase the number of strategy matrices from one to the number of stages and would increase the number of columns from two to three in all the matrices except the final one. Quotas, in turn, would impact on the conditional probabilities contained in the strategy matrix. The following [actuarial] selection strategies are typically distinguished [Gatewood & Feild, 1994; Milkovich & Boudreau, 1994]:

- multiple regression strategy [single-stage, compensation allowed];
- multiple cutoff strategy [single stage, limited compensation];
- multiple hurdle strategy [multi-stage, no compensation];
- profile comparison strategy [single stage, no compensation]; and
- hybrid strategies.

A [multiple] regression strategy was implicitly assumed in the preceding argument and will henceforth be formally assumed. A two-way strategy matrix is thus assumed. The information categories in such a strategy matrix would consist of class intervals formed on either the expected criterion score scale or the conditional probability of success scale. The nature of the cell entries in the strategy matrix would depend on whether a quota is in force and/or the aspiration level of the decision maker as it manifests

itself in the specification of a minimally acceptable criterion score and/or conditional probability for success.

Earlier it was argued that the obtained validity coefficient cannot, without appropriate corrections, be considered an unbiased estimate of the actual validity coefficient of interest. It was thus argued that the validity coefficient should be corrected for restriction of range, criterion unreliability, and the fitting of error by differential predictor weights before being transported to the area of application. If these corrections would be applied, the validity coefficient would be adjusted, but that still leaves the information contained in the strategy matrix unaffected. The decision strategy actually applied for decision-making is thus still the one derived from a simulated application, which, however, is not fully representative of the actual application. This, however, begs the question whether this matters; that is, do the characteristics in terms of which the context in which the decision strategy is actually applied differ from the context in which it was developed in any way affect the decision strategy, and if so, in which way[s]. If the considerations underlying the corrections previously applied to the correlation coefficient do in fact also affect one or more of the facets of the decision strategy, corrections to the decision rule would then be required. That would, however, in turn bring the following question to the fore again:

- would such corrections change decisions on employment applicants?; and if so
- what would the value of the consequences of the change in decisions be?; and
- how would the cost of the corrections compare to the value of the changed decisions affected by the corrections?.

Again, as previously, one could contend that such corrections applied to the decision rule do in fact serve no practical purpose if they do not change the actual decisions on applicants for employment or do so at a cost exceeding the value of the change in decisions.

It should be kept in mind that all the evidence required to develop and justify a selection procedure in terms of efficiency and equity is obtained in an artificial, simulated environment. The contention is thus, that the concern with external validity cannot logically be restricted to the validity coefficient only, but should extend throughout the whole argument in terms of which a selection procedure is developed and justified.

The all-important, but too often grossly neglected, question regarding the justifiability of a selection procedure shifts the attention from the development and operation of a selection strategy to the evaluation thereof. Evaluation of human resource interventions generally serve two purposes, namely [Rossi & Freeman, 1985; Weiss, 1972]:



- to determine the extent to which the intervention in question did serve the purpose it was designed for and in the manner intended, and in so doing, justify its current and continued usage; and
- to provide the diagnostic feedback needed to improve the intervention's ability to serve the purpose it was designed for.

According to Rossi and Freeman [1985, p. 19] evaluation research refers to:

the systematic application of social research procedures in assessing the conceptualization and design, implementation, and utility of social intervention programs. In other words, evaluation research involves the use of social research methodologies to judge and to improve the planning, monitoring, effectiveness, and efficiency of health, education, welfare, and other human service programs.

In all fairness to Rossi and Freeman [1985] it should be stated that they probably did not have selection procedures in mind when they referred to social intervention programs. Nonetheless, a wider interpretation of social intervention programs could legitimately include human resource selection procedures.

The rather comprehensive definition offered by Rossi and Freeman [1985] endorses the position enunciated earlier that the empirical investigation of the theoretical base of a selection procedure, so as to establish its validity, constitutes an important and necessary, but not sufficient, first phase of evaluation. A second phase of evaluation concerns the extent to which the actual decision-making practice departs from the instructions contained in the strategy matrix. If the decision maker should, contrary to the prescription of the selection strategy, differentiate between applicants falling in the same information category, based on additional clinical information, the chances are, despite strong personal convictions to the contrary, that the effectiveness and efficiency of his decision making would thereby be impaired [Cronbach & Gleser, 1965; Gatewood & Feild, 1994; Meehl, 1954; Tenopir, 1983]. The third phase of evaluation concerns the actual impact of the selection procedure. All three phases of evaluation, and not only the impact evaluation phase, are necessary to construct a sound justification for a selection procedure and all three phases afford valuable feedback to improve on the performance of a selection intervention.

### 1.2.2 A Strategic Perspective on Personnel Selection

Shifting the focus to the evaluation of impact results in a concomitant shift in the perspective from which selection is conceptualised in a normative decision-theoretic framework. A selection strategy

serves the purpose of guiding numerous decisions about applicants made by human resource managers and therefore should be evaluated by "its total contribution when applied to a large number of decisions" [Cronbach & Gleser, 1965, p. 23]. Thus the focus should now be placed on strategic decisions about selection procedures rather than on operational decisions about individual applicants [Boudreau, 1991; Cascio, 1991b]. The decision options under consideration are therefore not individual applicants, but rather different possible procedures or strategies that could be used to assign applicants to treatments [Boudreau, 1991; Cronbach & Gleser, 1965].

Multi-attribute utility [MAU] theory represent a decision theoretic approach directed at an analysis of the usefulness and desirability of decision options under conditions of uncertainty, multiple conflicting objectives, and costs and benefits accruing to various stakeholders [Huber, 1980; Keeny & Raiffa, 1976]. As a specific exponent of the latter approach, selection utility models are characterised by the following structural elements [Boudreau, 1991; Boudreau, 1989]:

- a set of decision options that represent the alternative procedures or strategies under consideration;
- a set of attributes affected by the decision, reflecting the characteristics of the outcomes considered relevant by the decision maker and pertinent constituencies;
- a utility or payoff scale reflecting the value of each attribute to the decision maker;
- a payoff function, reflecting the weight assigned to each attribute and the rule in terms of which the differently valued attributes are combined to estimate the total utility of each option under consideration; and
- a set of parameters characterising the decision situation in which the selection strategy will be used.

The option set most often considered by selection utility models comprises the use of a specific battery of selection instruments under the directive of a specific selection strategy versus random selection [Boudreau, 1991; Taylor & Russell, 1939]. Defining random selection as the baseline implies the value of selection procedures to be adequately captured by the extent to which it improves selection efficiency over and above, that which would have been achieved by chance alone. According to Cronbach and Gleser [1965] such reasoning can easily lead to overoptimistic estimates of selection utility if proper care is not taken in delineating the appropriate decision alternative with which to compare the selection procedure in question. The important point to recognise is that if the application of the selection procedure were impossible the decision maker need not fall back on chance decision making, but would base his decisions on whatever prior information is available. The decision option with which the envisaged selection procedure should be compared is thus the best a priori strategy [Cronbach & Gleser, 1965]. How this comparison should be made, and simultaneously, as was indicated earlier, the meaning afforded to the term "applicant population", depends on the way the new

selection procedure is positioned relative to the selection requirements already in use. Cronbach and Gleser [1965] identify the following three possibilities:

- all selection requirements currently employed will continue to be used for pre-screening, and thus the envisaged strategy following on pre-screening should be compared to the existing strategy [applicant population represented by pre-screened applicants];
- the new selection procedure will replace all current selection requirements, and thus the two strategies should be compared [applicant population represented by unscreened applicants]; or
- the new procedure and current selection requirements will be integrated to produce information for selection decision making, and thus the combined strategy should be compared to the existing strategy [applicant population represented by unscreened applicants].

Given a set of decision options, selection utility models specify an array of attributes in terms of which the outcomes generated by any option would be described [Boudreau, 1991; Keeny & Raiffa, 1976]. Cronbach and Gleser [1965, p. 22] point to the need for multiple descriptors when characterising the attribute domain by supplying the following, somewhat vague, definition:

all the consequences of a given decision that concerns the person making the decision [or the institution he represents].

An exhaustive description of all the possible consequences resulting from a selection decision would be so voluminous as to render any selection utility model useless as a practical decision aid. Thus, like any other type of model, selection utility models have to simplify reality by omitting some lesser important attributes. The attributes that should be retained are those most closely aligned with the decision objective[s]. Milkovich & Boudreau [1994] identify two general categories of human resource objectives, namely efficiency and equity, that could serve as beacons to guide the nomination of attributes for inclusion in the salient attribute set. Efficiency related attributes are those outcome characteristics that affect the organisation's ability to maximise output while minimising inputs. Equity related attributes, in contrast, are those outcome characteristics that affect the fairness of the selection procedure [Boudreau, 1991; Milkovich & Boudreau, 1994]. A small set of efficiency related attributes seem to have dominated selection utility conceptualisations. Boudreau [1991, p. 628] contends that at least three basic attributes should be considered to ensure an adequate selection utility model:

- "Quantity - the number of employees and time periods affected by the consequences of program options
- Quality - the average effect of the program options on work force value, on a per-person, per time-period basis

- Cost - the resources required to implement and maintain the program option"

The third component of a selection utility model is a set of utility or payoff scales for each attribute, established with the purpose of quantifying the level of each attribute, so as to articulate the decision maker's satisfaction with the various attributes of the outcomes that could possibly be obtained. Some abstract scale of utility might be employed for this purpose on which decision-makers subjectively scale their evaluation of the different outcome attributes. This would, however, ignore the need to translate the consequences of selection procedures into the everyday financial language of line management [Cascio, 1991b; Cronshaw & Alexander, 1985; Gow, 1985; Sheppeck & Cohen, 1985]. Consequently the quantity attribute is usually expressed in person-years while the cost attribute is usually measured in an appropriate monetary unit [Boudreau, 1989]. In the case of the quality attribute, general consensus exists that the appropriate scaling unit for this attribute should be dollars [that is, rand-cent] per person-year, although considerable disagreement still exists on what these values actually should reflect. Some form of clarity, however, emerges if it is argued that human resource selection procedures provide economic benefit to organisations through their ability to enhance the quality of the work force employed [Boudreau, 1991]. The increase in work force quality, in turn, manifests itself in increased work performance, which in turn manifests itself in improved output [Binning & Barrett, 1989; Campbell, 1991]. How the organisation capitalises on such quality enhancement depends on how it affects performance and eventually output. Both performance and output are multidimensional concepts, and improvements in the quality of employees can manifest itself in increases in performance and output in any one dimension, or combination of dimensions. At least three inter-related output dimensions should be taken into account, namely [Boudreau, 1989; Boudreau, 1991]:

- quantity of production or output;
- quality of production or output; and
- production cost.

This would then logically imply that organisations could derive economic benefit through enhanced labour force quality, brought about by improved human resource selection [or any other human resource intervention for that matter], by channeling the increased labour capability into either [Boudreau, 1991]:

- an increase in the quantity of production, and/or;
- an increase in the quality of production and/or;
- a reduction in production costs.

A utility scale for the quality attribute should thus be considered appropriately defined, if such a definition could reflect changes in any one, or combination, of the aforementioned output dimensions. At least three different interpretations of the payoff scale for the quality attribute seem to exist in selection utility applications [Boudreau, 1991]:

- payoff interpreted as cost reduction;
- payoff interpreted as increased value of output or revenue; and
- payoff interpreted as increased profits.



Only the profit interpretation of quality payoff encompasses all three output dimensions. This should, however, not lead to the conclusion that the other two interpretations are totally without merit. The applicability of the other two interpretations depends on the manner in which the organisation uses quality improvements brought about by improved selection strategies. Furthermore, a revenue based interpretation of quality utility could still in the end deliver a profit based total utility estimate for the option under consideration, by incorporating the cost output dimension in the payoff function [Boudreau, 1991]. The measurement of utility in general, but particularly the utility associated with the quality attribute, represents the Achilles' heel of selection utility analysis theory [Cronbach & Gleser, 1965]. This controversial theme will be discussed subsequently.

The fourth component of a selection utility model is the payoff function reflecting the weight assigned to each attribute and the rule in terms of which the differently valued attributes are combined to estimate the total utility of each option under consideration [Boudreau, 1989; Boudreau, 1991]. The payoff function in essence creates the common value base needed to compare options in terms of their ability to realize the decision objective. Such a common value base should, however, be related to the decision objective and should be expressed in the same metric as the objective [Lumby, 1994]. The payoff function applicable to selection utility models, should thus, given the fundamental position enunciated in the introductory section of this chapter, be derived from the primary organisational objective of profitability maximisation. Therefore, a specific selection strategy should be evaluated in its totality by its contribution to profit earned over a particular period relative to the capital used to generate that profit. To justify the investment of capital in a selection procedure in terms of the quest to add value to the organisation, the expected cash inflow, appropriately discounted, generated by the procedure, should therefore exceed the financial resources required to start-up and run the intervention. Boudreau [1991, p. 629] proposes the following basic structure:

The payoff function may be considered a variant of the cost-volume-profit models used in other managerial decisions to invest resources. The utility of an HRM program option is derived by subtracting cost from the product of quantity times quality, with the program exhibiting the largest positive difference being preferred.

To restrict the payoff function to this basic form would, however, leave the decision-maker with insufficient evidence to adequately evaluate and compare selection strategy alternatives. The fundamental problem with the basic quality times quantity minus cost conception of overall payoff lies in its disregard of the following relevant considerations [Lumby, 1994; Cronshaw & Alexander, 1985; Boudreau, 1991; Levy & Sarnat, 1994]:

- the time value of capital;
- taxation;
- variable costs;
- multiple cohort effects; and



➤ risk.

Thus the basic payoff function should be extended to reflect these considerations [See chapter 7].

The fifth and final component of a selection utility model is a set of parameters characterising the situation in which the selection strategy is applied. The parameters of interest are those describing the marginal criterion and composite predictor distributions for the applicant population. Of specific interest is the proportion of applicants falling above the hiring cutoff on the composite predictor, or the selection ratio [SR], and the proportion of applicants who would be successful if selection would occur without the envisaged selection strategy, or the base rate [BR] [Milkovich & Boudreau, 1994; Taylor & Russell, 1939]. In the case of restricted selection [i.e. a selection quota is in force], the selection ratio is determined by the number of vacancies relative to the number of applicants. In the case of unrestricted selection, however, the critical predictor cutoff, the mean and the standard deviation of the composite predictor distribution determine the selection ratio. The base rate is a function of the critical criterion cutoff defining minimum acceptable performance, the mean and the standard deviation of the marginal criterion distribution. These parameters are influenced by an intricate interaction between the aspiration level of the decision maker, the long-term effect of the selection procedure, recruitment strategies and the dynamics of the labour market [Boudreau, 1991].

Selection utility models serve as reality simplifying conceptual frameworks designed as aids for "describing, predicting and explaining the usefulness or desirability" of selection decision strategies, "and analysing how that information can be used in decision making" [Boudreau, 1989, p. 228] aimed at improving selection strategies. When viewed from a historical perspective, the evolution of selection utility models present a fairly systematic progression from somewhat unsophisticated models to detailed, complex and rather daunting contemporary models [Boudreau, 1991; Rauschenberger & Schmidt, 1987]. The following utility models can be differentiated in terms of their interpretation of utility/payoff:

- payoff defined in terms of the validity coefficient;
- payoff defined in terms of the success ratio;
- payoff defined in terms of expected standardised criterion performance;
- payoff defined in terms of a monetary valued criterion

Chapter 7 extends the preceding discussion by presenting a detailed analysis of the aforementioned utility models.

Earlier it was argued that the obtained validity coefficient cannot, without appropriate corrections, be considered an unbiased estimate of the actual validity coefficient of interest. It was, therefore, argued that the validity coefficient should be corrected for restriction of range, criterion unreliability, and the fitting of error by differential predictor weights before being transported to the area of application.

Furthermore it was contended that the concern with external validity cannot logically be restricted to the validity coefficient only, but should extend throughout the whole argument in terms of which a selection procedure is developed and justified. It was thus argued that if these corrections would be applied to the validity coefficient, the effect thereof on the selection strategy should therefore also be examined. The same logic, however, would also apply to the different interpretations of the utility concept. The question is thus firstly whether the statistical corrections for the attenuating effect of criterion unreliability and/or various forms of restriction of range would affect the different utility estimates in terms of expected value and/or estimate variability. Furthermore, should such corrections have any effect on the utility estimates, a second question emerges, namely whether the expected utility values and/or the risk associated with a specific selection option change to such an extent that the decision on the selection option is altered. The foregoing argument thus implicitly concedes that the value of statistical corrections of the validity coefficient in the final analysis depends on [Boudreau, 1984; Boudreau, 1991]:

- the probability that the corrections can change decisions on selection options;
- the consequences of the changed decisions; and
- the cost of applying the statistical corrections.

If the probability of such corrections changing the pre-correction choice of a selection option is small and/or the value of the change in selection option is less than the cost of applying the statistical corrections, all other possible effects momentarily ignored, little justification would actually exist for these corrections from a practical perspective.

### 1.2.3 An Equity Perspective on Personnel Selection

Should the human resource function be summoned, either by management or regulatory agencies, to defend its selection procedures, it would have to construct a reasoned justification based on the efficiency and equity of the selection procedure [Guion, 1991; Milkovich & Boudreau, 1994]. The different aspects that should be incorporated in the efficiency component required by such a justification were examined in the preceding discussion. The equity component required in such a defense has, however, been largely ignored up to now.

Human resource selection procedures represents a powerful instrument enabling the human resource function to add value to the organisation by virtue of its ability to regulate the quality and quantity of employees flowing into, through and out of the organisation. Human resource selection procedures derive their ability to add value to the organisation from their capability to discriminate between applicants in terms of attributes relevant to job performance. Selection measures are designed to

discriminate and in order to accomplish its professed objective it must do so [Cascio, 1991a]. However, due to the relative visibility of the selection mechanism's regulatory effect on the access to employment opportunities, the question readily arises whether the selection strategy discriminates fairly. Selection fairness, however, represents an exceedingly elusive concept to pin down with a definitive constitutive definition. The problem is firstly that the concept cannot be adequately defined purely in terms of psychometric considerations without any attention to moral/ethical considerations. The inescapable fact is that, due to differences in values, one man's foul is another man's fair. The problem is further complicated by the fact that a number of different definitions and models of fairness exist which differ in terms of their implicit ethical positions and which, under certain conditions, are contradictory in terms of their assessment of the fairness of a selection strategy and their recommendations on remedial action [Arvey & Faley, 1988; Cascio, 1991a ; Petersen & Novick, 1976]. The following three distinct fundamental ethical positions can be identified [Hunter & Schmidt, 1976]:

- unqualified individualism - selection strategies are ethically legitimate if they utilise all available means to maximise the correlation between  $Y$  and  $E[Y|X]$  and select those applicants with the highest expected criterion performance;
- qualified individualism; - selection strategies are ethically legitimate if they select those applicants with the highest expected performance but with the proviso that race, gender, creed or ethnicity shall not be considered even if these factors are related to performance [i.e. would have increased the correlation between  $Y$  and  $E[Y|X]$  had they been included in the selection battery]; and
- a quota position - selection strategies are ethically legitimate if the proportional composition of the labour market is reflected in the proportional composition of the work force selected.

A fairness model, based on one of the aforementioned ethical positions [or a variant thereof], serves the purpose of formalising the interpretation of the fairness concept. A fairness model thus permits the deduction of a formal investigative procedure needed to assess the fairness of a particular selection strategy should such a strategy be challenged to disprove a *prima facie* showing of adverse impact [Arvey & Faley, 1988]. A variety of fairness models have been proposed [Arvey & Faley, 1988; Cascio, 1991a; Petersen & Novick, 1976]. These fairness models examine the effect of the selection decision function on different subgroups contained in the applicant population by simulating the selection process on a representative sample from the applicant population. Assuming an applicant population consisting of two subgroups [ $\pi_1$  and  $\pi_2$ ] such a simulation would render information on two bivariate [normal] distributions characterised by the following parameters:

- a predictor mean [ $\mu[X]$ ], variance [ $\sigma^2[X]$ ], symmetry [ $\beta_1[X]$ ] and kurtosis [ $\beta_2[X]$ ];
- a criterion mean [ $\mu[Y]$ ], variance [ $\sigma^2[Y]$ ], symmetry [ $\beta_1[Y]$ ] and kurtosis [ $\beta_2[Y]$ ];
- predictor and criterion reliability coefficients [ $\rho_{ttX}$  and  $\rho_{ttY}$ ];
- a validity coefficient [ $\rho[X,Y]$ ];



- a regression equation  $E[Y_i | X] = \alpha + \beta X_i$  with a specific intercept  $[\alpha]$  and slope  $[\beta]$ ;
- a standard error of estimate  $[\sigma[Y | X]]$ ;
- a critical criterion cutoff  $[Y_c]$  and a critical predictor cutoff  $[X_c]$  and thus a specific base rate  $[BR=(a+c)/n]$ , selection ratio  $[SR=(a+b)/n]$  and a specific relationship between the number of true positives  $[a]$ , true negatives  $[d]$ , false positives  $[b]$  and false negatives  $[c]$ ; and
- utilities  $[U(O_j)]$  associated with the four possible selection outcomes  $[O_j]$ .

The various fairness models attempt to define unfairness in terms of differences on one or more of the aforementioned parameters across the subgroups  $[\pi_1]$  and  $[\pi_2]$  contained within the applicant population. At least thirteen different selection fairness models can be distinguished [Arvey & Faley, 1988; Cascio, 1991a; Petersen & Novick, 1976]. Only three of these models will be formally examined in this study, namely:

- the regression or Cleary model;
- the equal risk or Einhorn-Bass model; and
- the constant ratio or Thorndike model.

Chapter 7 extends the preceding discussion by presenting a detailed analysis of the aforementioned fairness models.

Earlier it was argued that the obtained validity coefficient cannot, without appropriate corrections, be considered an unbiased estimate of the actual validity coefficient of interest. It was, therefore, argued that the validity coefficient should be corrected for restriction of range, criterion unreliability, and the fitting of error by differential predictor weights before being transported to the area of application. Furthermore it was contended that the concern with external validity cannot logically be restricted to the validity coefficient only, but should extend throughout the whole argument in terms of which a selection procedure is developed and justified. It was thus argued that if these corrections would be applied to the validity coefficient, the effect of these corrections on the selection strategy, the different interpretations of the utility concept and the risk associated with different selection options should also be examined. The further question now arises whether the finding/verdict of the various preceding fairness models when applied to a specific selection strategy and their recommendations on remedial action, if required, would be affected by these corrections for criterion unreliability and/or restriction of range. Again one would argue as before that the value of statistical corrections of the validity coefficient in the final analysis depends on [Boudreau, 1984; Boudreau, 1991]:

- the probability that the corrections can change decisions on the fairness of a selection strategy;
- the consequences of the changed assessments of the fairness of a selection strategy; and
- the cost of applying the statistical corrections.

If the probability of such corrections changing the verdict on the fairness of a selection strategy is small and/or the value of the effect of a changed assessment of the fairness of a selection strategy is less than the cost of applying the statistical corrections, all other possible effects momentarily ignored, little justification would actually exist for these corrections, although from a logical-theoretical perspective they might seem to be indispensable.

### 1.3. SUMMARY AND RESEARCH OBJECTIVES

Selection, as it is traditionally interpreted, represents a critical human resource intervention in any organisation in as far as it regulates the movement of employees into, through and out of the organisation. As such selection thus firstly represents a potentially powerful instrument through which the human resource function could add value to the organisation [Boudreau, 1983b; Boudreau & Berger, 1985a; Cascio, 1991b; Cronshaw & Alexander, 1985]. Selection secondly, however, also represents a relatively visible mechanism through which access to employment opportunities are regulated. Because of this latter aspect, selection, more than any other human resource intervention, had been singled out for intense scrutiny from the perspective of fairness and affirmative action [Arvey & Faley, 1988; Milkovich & Boudreau, 1994]. Two basic criteria are thus implied in terms of which selection procedures need to be evaluated, namely efficiency and equity [Milkovich & Boudreau, 1994]. The quest for efficient and equitable selection procedures requires periodic psychometric audits to provide the feedback needed to iterate the selection procedure to greater efficiency and to provide the evidence required to acquit the organisation should it be challenged in terms of anti-discriminatory legislation. According to the Guidelines for the validation and use of personnel selection procedures [Society for Industrial Psychology, 1992], the Principles for the validation and use of personnel selection procedures [Society for Industrial and Organisational Psychology, 1987] and the Kleiman and Faley [1985] review of selection litigation, such a psychometric audit of a selection procedure would require the human resource function to demonstrate that:

- the selection procedure has its foundation in a scientifically credible performance theory;
- the selection procedure constitutes a business necessity; and
- the manner in which the selection strategy combines applicant information can be considered fair.

The empirical evidence needed to meet the aforementioned burden of persuasion is acquired through a simulation of the actual selection procedure on a sample taken from the applicant population. Internal and external validity constitute two criteria in terms of which the credibility and convincingness of the

evidence produced by such a simulation would be evaluated. The following two crucial questions are thereby indicated.

- to what extent can the validation researcher be confident that the research evidence produced by the selection simulation corroborates the latent structure/nomological network postulated by the research hypothesis within the limits set by the specific conditions characterising the simulation?; and
- to what extent can the validation researcher be confident that the conclusions reached on the simulation will generalise or transport to the area of actual application?

The conditions under which selection procedures are typically simulated and those prevailing at the eventual use of a selection procedure normally differ to a sufficient extent to challenge the transportability of the validation research evidence. Statistical corrections to the validity coefficient [Gulliksen, 1950; Pearson, 1903; Thorndike, 1949] are generally available to ex post facto enhance external validity<sup>2</sup>. The remainder of the argument in terms of which a selection procedure is developed and justified could, however, also be biased by any discrepancy between the conditions under which the selection procedure is simulated and those prevailing at the eventual use of the selection procedure. Relatively little concern, however, seem to exist for the transportability of the decision function derived from the selection simulation and descriptions/assessments of selection decision utility and fairness. This seems to be a somewhat strange state of affairs. The external validity problems with validation designs are reasonably well documented [Barrett, Philips & Alexander, 1981; Cook, Campbell & Peracchio, 1992; Guion & Cranny, 1982; Sussmann & Roberson, 1986]. It is thus not as if the psychometric literature is unaware of the problem of generalising validation study research findings to the eventual area of application. The decision function is probably the pivot of the selection procedure because it firstly captures the underlying performance theory, but more importantly from a practical perspective, because it guides the actual accept and reject choices of applicants [i.e. it forms the basis of the strategy matrix]. Restricting the statistical corrections to the validity coefficient would leave the decision function unaltered even though it might also be distorted by the same factors affecting the validity coefficient. Basically the same logic also applies to the evaluation of the decision rule in terms of selection utility and fairness. Correcting only the validity coefficient would leave the "bottom-line" evaluation of the selection procedure unaltered. Restricting the statistical corrections to the validity coefficient basically means that practically speaking nothing really changes. The fundamental research objective is:

- to determine whether any discrepancy between the conditions under which the selection procedure is simulated and those prevailing at the eventual use of the

---

<sup>2</sup> If such corrections would affect the probability of rejecting the null hypothesis, statistical conclusion validity would also be affected.

selection procedure produces bias in estimates required to specify and justify the procedure;

- to delineate appropriate statistical corrections of the validity coefficient, decision rule and descriptions/assessments of selection decision utility and fairness, required to align the contexts of evaluation/validation and application; and
- to determine whether the corrections should be applied in validation research.

The a-priori position is taken that these statistical corrections should be applied if they correct systematic bias in the obtained result [i.e. validity coefficient, decision function or description of selection utility and fairness] and if they change decisions regarding:

- the validity of performance hypotheses; and/or
- the choice of applicants to select; and/or
- the appropriate selection strategy option; and/or
- the fairness of a particular selection strategy.

The argument is thus by implication that there is little merit in applying statistical corrections should they not change any part of the total case built by the validation research team in defense of the selection procedure even if they should rectify systematic bias in the obtained estimates. The following, more specific, research objectives could thus be formulated:

- to determine the individual and combined effect of corrections for attenuation due to the unreliability of the criterion only and restriction of range on the magnitude of the validity coefficient;
- to determine the individual and combined effect of corrections for attenuation due to the unreliability of the criterion only and restriction of range on the magnitude of the standard error of the validity coefficient;
- to determine the individual and combined effect of corrections for attenuation due to the unreliability of the criterion only and restriction of range on the empirically derived exceedence probabilities  $\alpha_s$  or achieved significance level];
- to determine the individual and combined effect of corrections for attenuation due to the unreliability of the criterion only and restriction of range on the a priori probabilities  $[1-\beta]$  or power of the tests of the significance for the validity coefficient;
- to determine the individual and combined effect of corrections for attenuation due to the unreliability of the criterion only and restriction of range on the parameters [intercept, slope and conditional criterion variance] of the linear regression of the criterion on the predictor;



- to determine the individual and combined effect of corrections for attenuation due to the unreliability of the criterion only and restriction of range on the magnitude of the expected criterion performance conditional on the level of predictor performance; and
- to determine the individual and combined effect of corrections for attenuation due to the unreliability of the criterion only and restriction of range on the magnitude of the probability of a substandard criterion performance conditional on the level of predictor performance.

#### 1.4. STRUCTURAL OUTLINE OF DISSERTATION

Due to the nature of the objectives of the study, the structure of the dissertation deviates somewhat from the conventional format. In contrast to the more familiar format, the methodology used to achieve the stated objective is not contained in a single, dedicated chapter. Due to the multi-faceted nature of the objectives of the study, the ideal of a systematic and coherent presentation is best achieved by introducing a description of the appropriate methodology at the point in the argument where it is actually applied. The outline of the structure presented below does, however, provide an introductory overview of the methodology of the study.

Chapter 1 identified the objectives of the study and established, through a reasoned argument, the necessity of the envisaged research.

Classical test and measurement theory is discussed in Chapter 2 to create a theoretical foundation for the derivation and discussion of the aforementioned statistical correction formula in Chapter 3. Expressions for a number of reliability coefficients identified in the literature are derived analytically. The derivations presented in Chapter 2 are, however, not essential for the derivation of the correction formula for attenuation and restriction of range since these equations are neutral to the specific type of reliability that appears in them. Neither do any of these derivations constitute a new contribution to the psychometric literature. Although an extensive and detailed review of reliability could probably have been avoided, it is nonetheless considered useful, if not essential, in establishing a comprehensive, self-contained understanding of the topic involved. The extensiveness of the review, furthermore, represents a commentary [and a confession] on the level of sophistication with which reliability theory is typically treated in South Africa.

Chapter 3 presents the analytic derivation of the appropriate formula to correct the validity coefficient for the attenuating effect of criterion unreliability and/or restriction of range. The effect of these

correction formula are then investigated through computational solution of the equations and the subsequent mapping of the values of the corrected validity coefficient over a space defined by the parameters affecting its reaction.

The sampling distributions and standard errors associated with the various corrected validity coefficients are examined in Chapter 4. Appropriate expressions for the various standard errors are identified in the literature. The reaction of the standard error associated with the various corrected validity coefficients to changes in the parameters affecting its magnitude are examined by substituting ranges of values in the standard error equations and solving the equations. The effect of disattenuating the validity coefficient on the standard error is explored by mapping the ratio of the corrected to the uncorrected standard error over a space defined by the relevant parameters affecting their magnitude.

The effect of the corrections for restriction of range and/or criterion unreliability on statistical significance-testing are examined in Chapter 5. The question whether corrections for restriction of range and/or criterion unreliability affect the significance of the validity coefficient is answered by examining the ratio of the change in the value of the correlation to the change in the standard error of the correlation [i.e.  $(\rho^{\wedge}_c - \rho^{\wedge}) / (\sigma[\rho^{\wedge}_c] - \sigma[\rho^{\wedge}])$ ] relative to the ratio  $\rho^{\wedge} / \sigma[\rho^{\wedge}]$ . Chapter 5 argues that corrections to the validity coefficient would be important if they change decisions on the validity of hypotheses explaining variance in performance at a cost substantially lower than the value of the altered decision. In operational terms the critical question consequently is under which conditions the change in  $\alpha_b$  [both increases and decreases] produces movement past the critical value  $\alpha_c$ . The relevant parameters that need to be considered when defining the aforementioned conditions would be those contained in the expression for the standard error of the corrected correlation coefficient and those affecting  $\alpha_c$ . The behaviour of a derivation of the aforementioned ratio [G] under different values of the relevant parameters is examined through computer generated plots of G against those parameters. Such a plot would indicate whether corrections to the correlation coefficient affect  $\alpha_b$ , how they affect  $\alpha_b$  and under what conditions. Such a plot would, however, be incapable of indicating whether the change in standard score [or  $\alpha_b$ ] is sufficient to affect the decision on  $H_0$ . This shortcoming is, however, be circumvented by plotting G on a continuous and a discrete/nominal scale simultaneously through the utilisation of appropriate character codes.

Chapter 6 examines the effect of restriction of range and criterion unreliability on the parameters of the regression of the criterion on the predictor. The effect of the statistical corrections on the slope- and intercept-parameters are examined analytically by algebraically deriving expressions for the disattenuated regression of the criterion on the predictor. A similar analytic approach is used to derive formula for the disattenuated conditional criterion standard deviation or standard error of estimate. In the case of Case 3[i] [Case C[i]] restriction of range, additional numerical simulations and graphical displays are used to examine how the analytically derived corrections affect the slope parameter and the



## CHAPTER 2

### MEASUREMENT THEORY

The purpose of the ensuing chapter is to systematically unfold a formal explication of the classical measurement theory so as to provide a theoretical model in terms of which the derivation of the previously mentioned correction formula can be formally demonstrated and comprehensively discussed. In essence it is argued that the core of a selection procedure is the performance theory/hypothesis on which the selection procedure is based. The performance hypothesis is a set of interrelated constructs, their definitions and propositions about the presumed interrelationship between them. Justification of a selection procedure requires empirical confirmation of the validity of propositions contained in the performance hypothesis. To empirically investigate the validity of such a performance hypothesis, however, requires information on the constructs comprising the hypothesis. Due to the abstract character of constructs, information on their state or level can be estimated only from their effects. Thus measurement of constructs by necessity is of an indirect, inferential nature, through the observation of indicants of the construct assumed or demonstrated to be related to the property. The acceptance of operationalism as a solution to the measurement problem posed by the abstract nature of psychological constructs rests on the supposition that inter- and intra-individual variance in the observed indicant can be explained solely in terms of differences in the underlying construct of interest. Although this represents a practically unattainable ideal, it nevertheless defines the objective of perfectly controlled measurement in which all extraneous variables are controlled. The unattainability of perfectly controlled measurement implies that any observed measure unavoidably, to some extent, is contaminated by measurement error. If the fallible observed score contains an error of measurement, an inescapable counterpart is thereby implied, namely an infallible measure without error or true score [Feldt & Brennan, 1989; Lord & Novick, 1968]. Any attempt to explicate the meaning of these measurement component terms, and to derive from their definitions suitable indicators of measurement quality, necessitates a formal measurement theory or model. Lord and Novick [1968, pp. 13-14] express the need for a measurement theory as follows:

One reason we need to have a theory of mental testing is that mental test scores contain sizable errors of measurement. Another reason is that abilities or traits that psychologists wish to study are usually not directly measurable; rather they must be studied indirectly, through measurement of other qualities. .... One function of any measurement theory employing hypothetical constructs and measurement deviations from these hypothetical values is the explanation of paradoxical events.



## 2.1 CONSTRUCTS

Nature would, in the absence of concepts/constructs, be experienced as a cacophonous, buzzing, whirling mess of sensations. Comprehension, contemplation and communication under these conditions would be almost impossible. A construct is an "in the head variable" [Kerlinger, 1986], an intellectual construction of the mind [Guion, 1991; Margenau, 1950], a cognitive building block, a link between Nature and reason [Margenau, 1950], a cognitive category created to enable man to intellectually organise/categorise the sensory confusion, to obtain an intellectual grip on that which he observes around him and to communicate such an understanding to his fellow man [Mouton & Marais, 1985]. Constructs do not arise naturally out of man's sensory experience of Nature; they are not wholly determined by perception [Margenau, 1950]. Constructs are words representing abstract ideas [i.e. concepts] created/constructed by man to be used to understand and explain phenomena in nature. Although the terms concept and construct have similar meaning they should not be regarded as synonyms. A concept represents an abstraction formed by generalizing a common/shared theme contained in observable particulars [Kerlinger, 1986]. The same is true for a construct. A construct "has the added meaning, however, of having been deliberately and consciously invented or adopted for a special scientific purpose" [Kerlinger, 1986, p. 27]. Science has as its principle objectives the description of empirical phenomena in Nature and the establishment of general principles in the form of theories in terms of which the empirical phenomena in Nature can be explained and predicted [Kerlinger, 1950; Torgerson, 1950]. Scientific theory represents a set of interrelated constructs, their definitions and propositions on the relationship between constructs with the purpose of explaining and predicting empirical phenomena in Nature [Kerlinger, 1986, p. 9]. Constructs thus form the primary structural component from which science constructs explanatory systems or nomological networks.

To be acceptable to science, and to distinguish them from "shadowy" non-scientific concepts, constructs must satisfy two types of prerequisites [Margenau, 1950]. The first demand comprises a set of six metaphysical principles, which have to be adhered so as to ensure satisfactory explanatory systems. Margenau [1950] lists and describes the following metaphysical requirements on constructs:

- Logical fertility: "constructs shall be so formulated as to permit logical manipulations. ... the constructs shall obey logical laws" [Margenau, 1950, p. 82].
- Multiple connections: "constructs admissible in science must be multiply connected; they may not be insular or peninsular; sets forming an island universe must be excluded" [Margenau, 1950, p. 87].
- Permanence and stability: "the constructs generated in explanation of a set of immediate experiences must, so long as the theory of which they form a part is accepted, be used with utmost respect for their integrity of meaning in all applications" [Margenau, 1950, p. 90].

- Extensibility of constructs: “Constructs, we recall, enter into two types of relation: with Nature and with other constructs. Hence they should be extensible in these two ways” [Margenau, 1950, p. 93]. Constructs should be extensible in the sense that they develop numerous connections to nature and develop direct connections too hitherto only indirectly connected constructs.
- Causality: “constructs shall be chosen as to generate causal laws” [Margenau, 1950, p. 96]
- Simplicity and elegance: “When two theories present themselves as competent explanations of a given complex of sensory experience, science decides in favor of the ‘simpler’ one” [Margenau, 1950, p. 96].

The second demand put to scientific constructs is that they should be empirically verifiable. A theory could be considered valid [Margenau, 1950] or corroborated [Popper, 1972] if empirically testable predictions derived/deduced from the theory are confirmed in a sufficient number of instances [Margenau, 1950]. Theories are valid to the extent that they have survived a sufficient number of attempts to falsify them [Popper, 1972]. The constructs which collectively constitute a valid theory may then be termed valid scientific constructs or verifacts [Margenau, 1950, p. 105].

Conceptualisation and operationalisation represent the two mutually complementary processes through which the meaning of constructs are explicated. Two dimensions of meaning can be distinguished, namely [Mouton & Marais, 1985]:

- a connotative dimension; and
- a denotative dimension.

The connotative dimension refers to the internal structure of the intellectual idea represented by the construct, inferred from its position in a nomological network relative to other constructs and consequently its function in the network. The connotative meaning of a construct is explicated through a process of conceptualisation whereby a constitutive, literary, syntactic or theoretical definition [Kerlinger, 1986; Lord & Novick, 1968; Marais & Mouton, 1985; Margenau, 1950; Torgerson, 1958] is established to describe the nature or structure of the abstract idea represented by the construct. Constructs are constitutively defined in terms of other constructs contained in the theory [Kerlinger, 1986; Margenau, 1950; Torgerson, 1958]. Constitutive definitions of constructs are thus necessarily circular in nature [Torgerson, 1958]. Conceptualisation thus represents an attempt to attain an intellectual grasp on the construct. All constructs in a scientific theory should be conceptualized [Margenau, 1950; Torgerson, 1958] in terms of valid constitutive definitions. The conceptualisation of a construct could be considered connotatively or theoretically valid if [Mouton & Marais, 1985]:

- all dimensions of meaning, implied by the way the construct is used, are identified; and
- these dimensions of meaning are mutually exclusive.

To be considered a scientific theory a sufficient number of its constructs must be connected directly to empirical phenomena in Nature by rules of correspondence [Margenau, 1950; Torgerson, 1958] so as to permit empirical testing of the theory. The denotative dimension refers to the array of concrete phenomena [i.e. objects, events, behavioural acts] indicated by the construct as constitutively defined. The explication of the denotative meaning of a construct is thus contingent on the explication of the connotative meaning. The denotative meaning of a construct is explicated [Carnap, 1950] through a process of operationalisation whereby an operational or semantic or epistemic definition [Kerlinger, 1986; Lord & Novick, 1968; Marais & Mouton, 1985; Margenau, 1950; Torgerson, 1958;] is established. The operational definition describes the visible manifestation of the abstract idea represented by the construct or describes the actions through which the construct could be manipulated to different conditions so as to obtain an empirical grasp on the construct. An operational or epistemic definition establishes a rule of correspondence between a construct and Nature [Margenau, 1950, p. 235]. Two types of operational definitions can be distinguished, namely [Kerlinger, 1986]:

- measured operational definitions; and
- experimental operational definitions.

The latter type of operational definition spells out the operations or actions required to alter, through manipulation or force, the condition or level of the construct. The first, more prevalent, type of operational definition, in contrast, specifies the operations or actions required to elicit observable behavioural denotations in which the construct manifests itself. Operational definitions, through the explication of the denotations of a construct, permit the development of measurement procedures so as to accrue the benefits of quantification.

## 2.2 THE NATURE OF MEASUREMENT

Measurement, according to Jones [1971, pp. 336-337]:

... is a purposive acquisition of information about the object, organism, or event measured by the person doing the measuring. It is a determination of the magnitude of a specified attribute of the object, organism, or event in terms of a unit of measurement. The result of the measurement is expressed by a numeral. The classification of attributes, either qualitative or quantitative, is distinguished from the measurement of attributes, which must be quantitative.



Campbell [1956, pp.1797-1801] interprets measurement as:

... the assignment of numbers to represent properties. .... The first point I want to notice is that it is only some properties and not all that can be thus represented by numbers. .... The measurable properties of a body are those which are changed by the combination of similar bodies; the non-measurable properties are those that are not changed. .... If a property is to be measurable it must be such that [1] two objects which are the same in respect of that property as some third object are the same as each other; [2] by adding objects successively we must be able to make a standard series one member of which will be the same in respect of the property as any other object we want to measure; [3] equals added to equals produce equal sums. In order to make a property measurable we must find some method of judging equality and of adding objects, such that these rules are true?

If the rather stringent requirements suggested by Campbell [1956] would be accepted, most if not all properties of interest to the behavioural sciences would have to be considered immeasurable. A significantly broadened interpretation of measurement suggested by Stevens [1946, p. 677] has greatly influenced subsequent thinking about measurement in the behavioural sciences:

Measurement in the broadest sense, is defined as the assignment of numerals to objects or events according to rules.

Although differing only slightly from the basic definition proposed by Campbell [1956], the consequence of the subtle change of wording is substantial. The critical difference between the first two definitions and the Stevens definition lies in the extent to which they require the numeric system and the to be measured property to be isomorphic to permit measurement.

Measurement is possible because, and to the extent that, there is a similarity between the characteristics of the numerical system on the one hand and the characteristics of the attribute to which these numbers are assigned on the other hand. The numerical system thus can be used as a model to describe and represent/replace the property to be measured. A model can be seen as a simplified, more familiar, "as if" representation of a phenomenon which is permissible because of, and to the extent to which there is, a similarity between the model and the phenomenon. Stevens [1946] argues that the characteristics of the numerical system always permits four possible hierarchically ordered operations [identification/classification, rank ordering, determining equality of differences and determining equality of ratio's] while the to be measured property permits one or more, but not necessarily all four, operations. The rule in terms of which numerals are assigned according to the Stevens definition could be based on any one of the permissible operations. Thus, according to Stevens [1946] four possible levels of similarity can exist between the numerical model and the to be measured property, thus resulting in four possible levels of measurement, namely nominal, ordinal, interval and ratio

measurement scales. The first two definitions, in contrast, restrict the term measurement to properties exhibiting the characteristic of additivity and thus the latter two of the aforementioned operations.

Rephrasing the Stevens [1946] definition so as to acknowledge the fact that a property of an individual is measured, and not the individual as such, and so as to acknowledge the abstract nature of the to be measured property, results in the following final definition of measurement:

Measurement is the assignment of numerals to an indicant of a property of an individual according to certain rules.

The following logic is thus implied. A psychological measurement procedure elicits a sample of behaviour through a sample of standardised stimuli under standardised condition. The stimulus sample is so constructed so as to reflect the underlying construct of interest through the testee's reaction to it, in that the quality or nature of the behavioural response to the stimuli is dependent on the underlying construct. Procedures are finally formulated in terms of which the elicited behaviour is observed, recorded and quantified. Given the [assumed] dependence of the behavioural response to the stimulus sample on the construct of which quantitative information is desired, differences in observed scores obtained by  $n$  testees should therefore indicate [on at least an ordinal scale] differences in the construct of interest. Reflecting the foregoing logic, the HSRC [Owen & Taljaard, 1988] defines a psychological test as:

... a purpose-specific evaluation and assessment procedure used to determine characteristics of people in areas of intellectual ability, aptitude, interests, personality profile and personality functioning. It consists of a collection of tasks, questions or items aimed at eliciting a certain type of behaviour under standard circumstances, from which scores with acceptable psychometric characteristics are inferred according to prescribed procedures.

The acceptability of the preceding argument is contingent on the validity of the premise that the behavioural response to the stimulus sample is in fact contingent on the construct of interest. Only then will inter- and intra-individual variance in the obtained/observed scores reflect only differences in the construct of interest. This central premise can, however, never be entirely attained. The behavioural response to the stimulus sample is never solely a function of the to be measured construct. The observed behavioural response is always also partially a function of other stable and systematic, though irrelevant attributes as well as an array of [unknown] unstable attributes whose combined influence exhibits a random-like character [although individually actually exerting a systematic influence]. Thus, to a certain degree, the observed measure contains measurement error [Feldt & Brennan, 1989; Lord & Novick, 1968; Stanley, 1971]. The acceptance of the fallibility of the observed score due to errors of measurement in turn, inescapable implies the concept of measurement without

error; therefore an infallible or true score [Feldt & Brennan, 1989; Lord & Novick, 1968]. The precise meaning of the true score component of measurement is dependent on the interpretation of measurement error; different definitions/interpretations of the true score concept are therefore possible. Two kinds of measurement error should, however, be distinguished, given the preceding argument, namely systematic error and random error [Guion, 1965]. The two kinds of measurement error should, in addition, be further subdivided in errors attributable to the measuring instrument or the measurement situation on the one hand and errors attributable to irrelevant attributes of the individual being systematically measured on the other. Systematic measurement error arising from the measuring instrument and/or the measurement situation is of relatively little concern since it does not affect inter- or intra-individual variance in the observed scores. Systematic measurement error arising from irrelevant [though stable] attributes of the testee, in contrast, has greater practical significance because it does produce irrelevant inter-individual variance although still no intra-individual variance. Random measurement error, whether arising from the measuring instrument, the measurement situation or attributes of the testee, produces inter- and intra-individual variance. Any extraneous influence, not relevant to the purpose of measurement, that produces variance in the observed scores over and above that produced by the construct of interest thus creates ambiguity in the meaning of the variance in the observed scores. Even if only theoretically attainable, the aforementioned premise nevertheless constitutes the theoretical ideal of controlling all extraneous variables which could affect the observed test score so that the variance in the obtained test results can be explained only in terms of the variance in the underlying construct of interest.

Control is pursued through two processes aimed at either removing the irrelevant variables or keeping the irrelevant variable constant; the effect of both being that the variables no longer produce variance in observed scores. The two processes in question are standardisation and test construction/item analysis [Crocker & Algina, 1986; Ghiselli, Campbell & Zedeck, 1981]. Standardisation is an attempt to control stimulus and evaluation/scoring related variables by attempting to keep these variables constant over different times, places and users. Item analysis, in contrast, attempts to control non-relevant constructs by trying to deprive them of the opportunity to influence test behaviour through the elimination of inappropriate items/stimuli from the stimulus set. Perfect control, however, is never achieved. The question, consequently, arises to what extent these processes did succeed in controlling extraneous variance. The following descriptive terms depict, each from a different perspective, the extent to which standardisation and item analysis did succeed in their endeavor to control extraneous variables [i.e. did succeed in minimising measurement error]:

- objectivity;
- reliability; and
- validity.

In as far as these descriptive terms reflect the extent to which obtained measures are contaminated by measurement error, the constitutive and operational definition of these terms, especially in the case of



the latter two descriptive concepts, would require a clear explication of the concepts of fallible observed score, measurement error and the inescapable counterpart of measurement error, infallible measurement without error. This in turn would require some form of representation [i.e. a formal model] of the measurement process and the resultant measures so as to establish the necessary cognitive system in terms of which these concepts can be defined.

### 2.3. MEASUREMENT THEORY

Measurement theories, or probably more accurately measurement models, constitute simplified "as if" representations of the measurement process and the resultant measures. Measurement models function as conceptual and analytic aids in understanding measurement and in deriving psychometric procedures/indicators to estimate the error in measurement and to affect various corrections. A measurement model thus represents a useful way of thinking about measurement which does not claim to portray the actual structure and dynamics of measurement but which permits the derivation of valid conclusions and useful psychometric indicators because of sufficient isomorphism between model and measurement. The classical measurement model refers to the historically oldest body of measurement assumptions and derived results. Two other models of more recent origin have since been formulated that present alternative conceptualisations of the measurement process. The domain sampling model, or generalisability theory as some of its more recent modifications have been labeled, represents one alternative to the classical position [Cronbach, Gleser, Nanda & Rajaratnam, 1972; Shavelson & Webb, 1991]. Latent trait or item response theory [Drasgow & Hulin, 1991; Hambleton, 1989; Lord & Novick, 1968] represents the third major theoretical position. Each measurement model has certain strengths and weaknesses associated with it [Crocker & Algina, 1986; Hambleton, 1989;]. The development of the latter two models flowed from the recognition of certain shortcomings in the classical model [e.g. the group dependence of the difficulty and the discrimination indexes] and the inappropriateness of the classical true score model to investigate specific problems and themes [e.g. tailored testing, item bias analysis and detection and test equating]. Although these models occasionally produce different results from those derived from the classical model, they for the most part yield similar and consistent results despite their differences in formulation [Feldt & Brennan, 1989; Ghiselli, Campbell & Zedeck, 1981]. Classical measurement theory developed, as did the domain sampling model, first and foremost to explain inconsistencies in measurement and thereby derive quantitative indexes of the reliability of a set of observed scores.

## 2.4 NOTATION

In the subsequent sections the classical approach to measurement theory will be presented and discussed. An appropriate notational system is needed to pursue this objective. The conventional Greek symbols will be used to represent population parameters:  $\sigma^2$  for variance,  $\mu$  for mean,  $\rho$  for correlation. Sample estimators of population parameter are indicated by modifying the parameter symbol by placing a caret [^] immediately to the right of it. Parameters and parameter estimates will carry suitable qualifiers to identify the variables involved. The following notation, based on the notational system used by Lord and Novick [1968], will be used;  $\sigma^2[X]$ ,  $\mu[X]$ ,  $\rho[X,Y]$ ,  $\sigma[X,Y]$  and  $\beta[X,Y]$ . The symbol  $\sigma[X,Y]$  will denote co-variance and the standard error of estimate associated with the regression of Y on X will be denoted as  $\sigma[Y|X]$ . The reliability coefficient for the predictor and criterion measures will be denoted  $\rho_{ttX}$  and  $\rho_{ttY}$  respectively. Lowercase qualifiers and subscripts [i.e.  $\rho[x,y]$ ,  $\sigma[y]$  or  $\rho_{ttx}$ ] will be used to indicate that the parameters apply to a restricted population whereas capital letter qualifiers and subscripts will denote parameters of an unrestricted population. Capital letters [X, Y, Z, T & E] are used to denote random variables. Lower case Greek letters [ $\xi$ ,  $\psi$ ,  $\zeta$ ,  $\tau$ , &  $\epsilon$ ] are used to indicate sample observations on the corresponding random variables. Random variables and observations will carry suitable subscripts to indicate the testee and test form involved. The Greek capital letter epsilon [E] will be reserved for use as the expectation symbol. . X and Y will denote the observed predictor and criterion variables respectively.  $T_X$ ,  $T_Y$ ,  $E_X$  and  $E_Y$  will denote the true and error score components of the unrestricted observed predictor and criterion variables. The true and error score components of the restricted observed predictor and criterion variables will be denoted by corresponding lowercase letters and subscripts. The Greek capital letter pi [ $\Pi$ ] will be used to represent a population whereas the lowercase letter pi [ $\pi$ ] with suitable subscripts will denote groups within  $\Pi$ .

Additional symbols needed for subsequent derivations will be introduced and defined as and when required. It is unavoidable that the notational system used in this dissertation will differ from the various notational systems employed in other studies on the same or similar topics. Some form of translation from one representational system to another might therefore be required when comparing the results reported in this dissertation with results reported elsewhere. In addition, it is unavoidable that occasions will arise where the same basic symbol is employed for different purposes. This is a problem generally encountered in statistical and psychometric texts and not a problem unique to this study. The specific interpretation of these multi-meaning symbols will be explicitly defined in each particular context. Care should consequently be taken not to extrapolate the meaning of a symbol beyond the context in which it was defined.



## 2.5 CLASSICAL MEASUREMENT THEORY

The origin of the classical measurement model can be traced back to British psychologist Charles Spearman [1904; 1907; 1913]. Subsequent contributions by Guilford [1936], Gulliksen [1950], Lord and Novick [1968], Magnusson [1967] and Thorndike [1949], amongst others, elaborated on the foundations laid by Spearman to produce the form of the classical measurement model as it is presented here.

The initial assumption of the classical measurement model is that measurement occurs with respect to a specific trait/construct which remains stable over the period of testing and that successive measurements with parallel test forms are unaffected by previous measurements. Consequently, if variations in the observed test scores obtained by any testee over the parallel test forms would be found, such variation would be view the result of measurement error. The essence of the classical measurement model is that any observed score  $X_{ij}$ , obtained by [a randomly selected] testee  $i$  in a [denumerably infinite] population of  $n$  [ $n \rightarrow \infty$ ] testees on test form  $j$  selected from  $k$  alternative parallel tests, represents a random variable conceptualised as a linear composite of two hypothetical components - a true score component  $[T]$  and an error score component  $[E]$  - as shown in Equation 2.1.

$$X_{ij} = T_{ij} + E_{ij}; i=1, 2, \dots, n \quad j=1, 2, \dots, k \text{-----} 2.1$$

The observed score  $\xi_{ij}$  a testee  $i$  actually obtains on any single administration of any single test form  $j$  is conceived to be the realisation of a random process. The score  $\xi_{ij}$  a testee  $i$  actually obtains on any single administration of any single test form  $j$  thus represents a random sample of one outcome from a hypothetical distribution of possible outcomes [Lord & Novick, 1968; Novick, 1966].

Each score  $\xi_{ij}$  a testee  $i$  actually obtains on any single administration of any single test form  $j$  represents a linear combination of a realisation  $\tau_{ij}$  of the random variable  $T_{ij}$  and a realisation  $\epsilon_{ij}$  of the random variable  $E_{ij}$  [Crocker & Algina, 1986; Feldt & Brennan, 1989; Gulliksen, 1950; Lord & Novick, 1968; Novick, 1966]. For any fixed testee  $i$  the true score is constant but the observed score and error score remain random variables. Equation 2.1 thus implies a  $[n \times k]$  matrix of observed test scores  $\xi_{ij}$  obtained by  $n$  testees on  $k$  parallel measurements as shown in Table 2.1.

According to Gulliksen [1950, p. 11] two tests [or more generally, measures] may be said to be parallel when "it makes no difference which test you use". Parallel measures are measures constructed to the same specifications, measuring the same construct/trait to the same degree [i.e. they evoke the same psychological processes]. Parallel measures thus give rise to identical distributions of observed scores

for any population approaching infinity [i.e. they have equal means and standard deviations/variances] which correlate equally with each other and correlate equally with any external variable [Feldt & Brennan, 1989; Ghiselli, Campbell & Zedeck, 1981; Gulliksen, 1950; Lord & Novick, 1968]. Parallel measures could alternatively be defined simply as separate, distinct but identical measures. The columns in Table 2.1 therefore have equal means and variances and the correlation between all pairs of columns are the same.

Table 2.1: Observed scores on k parallel measurements by n testees

PARALLEL MEASUREMENTS			
	1	.. j ..	k
1	$\xi_{11} = \tau_{11} + \varepsilon_{11}$	$\xi_{1j} = \tau_{1j} + \varepsilon_{1j}$	$\xi_{1k} = \tau_{1k} + \varepsilon_{1k}$
2	$\xi_{21} = \tau_{21} + \varepsilon_{21}$	$\xi_{2j} = \tau_{2j} + \varepsilon_{2j}$	$\xi_{2k} = \tau_{2k} + \varepsilon_{2k}$
.			
.			
.			
i	$\xi_{i1} = \tau_{i1} + \varepsilon_{i1}$	$\xi_{ij} = \tau_{ij} + \varepsilon_{ij}$	$\xi_{ik} = \tau_{ik} + \varepsilon_{ik}$
.			
.			
.			
n	$\xi_{n1} = \tau_{n1} + \varepsilon_{n1}$	$\xi_{nj} = \tau_{nj} + \varepsilon_{nj}$	$\xi_{nk} = \tau_{nk} + \varepsilon_{nk}$

The term test and the term measurement are quite often and somewhat casually, used interchangeably as if synonymous in meaning. Psychological tests are composed of items grouped together in sub-tests. Each item, subtest or test generates a measurement. The subtest measurement is additively determined from the relevant individual item measures, while the total test measure is analogously determined from the subtest measurements. Each item measurement thus represents a component in a subtest composite measurement and each subtest measurement, in turn, represents a component in the total test measurement. The concept of parallel measurements thus could apply to measurements generated by individual items, sub-tests or tests.

The error score, when used in discussions on reliability of measurement, is considered to be essentially random and therefore unpredictable with regards to size and direction. Thus, by assumption the expected value of the random variable  $E_{ij}$  taken over forms [i.e. for any row in Table 2.1] and taken over testees [i.e. for any column in Table 2.1] is zero as shown by Equation 2.2 [Crocker & Algina, 1986; Ghiselli, Campbell & Zedeck, 1981; Gulliksen, 1950; Thorndike, 1982].

$E[E_{ij}] = 0; i=1, 2, ..., n; j=1, 2, ..., k$  ----- 2.2

Furthermore it is assumed that the corresponding true and error scores are uncorrelated [Equation 2.3], that error scores on different parallel measures  $p$  and  $q$  are uncorrelated [Equation 2.4] and that error scores are uncorrelated with any external/criterion variable [Equation 2.5] [Ghiselli, Campbell & Zedeck, 1981; Gulliksen, 1950; Thorndike, 1982].

$$\rho[T, E] = 0; i=1, 2, \dots, n; j \text{ ----- } 2.3$$

$$\rho[E_p, E_q] = 0; i=1, 2, \dots, n; j \text{ ----- } 2.4$$

$$\rho[E, Y] = 0; i=1, 2, \dots, n; j \text{ ----- } 2.5$$

These assumptions [Equations 2.2 – 2.5] collectively serve as a formal definition of the measurement error/error score concept as interpreted by the classical measurement model.

From the preceding assumptions of the classical model, Equation 2.6 is obtained.

$$\begin{aligned} E[T_{ij}] &= E[X_{ij} - E_{ij}]; i=1, 2, \dots, n; j \\ &= E[X_{ij}] - E[E_{ij}] \\ &= E[X_{ij}] \text{ ----- } 2.6 \end{aligned}$$

Since Equation 2.2 assumes that  $E[E_{ij}] = 0$

The expected true score is equal to the expected observed score; both expectations taken over testees. From Equations 2.1 and 2.2 it follows that the true score  $\tau_{ij}$  of testee  $i$  on measurement  $j$  is defined as the expected value of the observed score for testee  $i$  taken over forms, as shown by Equation 2.7

$$\begin{aligned} E[X_{ij}] &= E[T_{ij} + E_{ij}]; j=1, 2, \dots, k; i \\ &= E[T_{ij}] - E[E_{ij}] \\ &= \tau_{ij} - 0 \\ &= \tau_{ij} \text{ ----- } 2.7 \end{aligned}$$

Since  $T_{ij}$  is a constant for any  $i$  over all values of  $j$  and Equation 2.2 assumes that  $E[E_{ij}] = 0$

Thus defined, the true score does not [necessarily] represent a pure and comprehensive numerical reflection of the construct of interest. To a certain extent the term true score is therefore a bit of a misnomer. Equation 2.1 considers  $T_{ij}$  to be a reflection of all systematic influences shaping the testee's behavioural response to the stimulus sample;  $T_{ij}$  thus represents a composite of a pure and comprehensive reflection of the stable construct of interest [a systematic, relevant influence] and

systematic, stable but non-relevant influences. The error term  $E_{ij}$  represents only the residual error which is random and unpredictable [Guion, 1965; Stanley, 1971]. Taken over test forms  $\tau_{ij}$  is a constant [i.e.  $\sigma_j = 0$ ]. Taken over testees, however  $0 < \tau_{ij} < \infty$ . Variability in observed test scores over  $k$  parallel test forms for any testee  $i$  is therefore attributable to the random error term  $E_{ij}$ ; the larger the contribution of  $E_{ij}$  to the observed score  $X_{ij}$ , the lower the consistency/reliability of the observed measures. Thus for an array of observed scores  $X_{ij}$ , obtained over  $k$  parallel test forms for a testee  $i$ , all the observed test variance  $\sigma^2[X]$  is random error variance  $\sigma^2[E]$  as shown by Equation 2.8.

$$\begin{aligned}\sigma^2[X_{ij}] &= \sigma^2[T_{ij} + E_{ij}]; j=1, 2, \dots, k; i \\ &= \sigma^2[T] + 2\sigma[T,E] + \sigma^2[E] \text{-----} 2.8.1\end{aligned}$$

However, for any testee  $i$ :

$$\sigma^2[T] = 0 \text{-----} 2.8.2$$

and:

$$\sigma[T,E] = \rho[T,E]\sigma[T]\sigma[E] = 0 \text{-----} 2.8.3$$

Since Equation 2.3 assumes that  $\rho[T,E] = 0$  and since  $\sigma^2[T] = 0$

Therefore it follows that for any testee  $i$ :

$$\sigma^2[X] = \sigma^2[E] \text{-----} 2.8$$

By contrast, an array of observed scores  $X_{ij}$ , obtained for  $n$  testees on a single test form  $j$ , the observed score variance  $\sigma^2[X]$  over  $n$  testees is equal to the sum of the true score variance  $\sigma^2[T]$  and the random error variance  $\sigma^2[E]$  as shown by Equation 2.9.

$$\begin{aligned}\sigma^2[X] &= \sigma^2[T_{ij} + E_{ij}]; i=1, 2, \dots, n; j \\ &= \sigma^2[T] + 2\sigma[T,E] + \sigma^2[E] \text{-----} 2.9.1\end{aligned}$$

However:

$$\sigma[T,E] = \rho[T,E]\sigma[T]\sigma[E] = 0 \text{-----} 2.9.2$$

Since Equation 2.3 assumes that  $\rho[T,E] = 0$

Therefore for any test form j:

$$\sigma^2[X] = \sigma^2[T] + \sigma^2[E] \text{-----} 2.9$$

If Equation 2.9 would subsequently be multiplied by  $1/\sigma[X]$ , Equation 2.10 results.

$$1 = \sigma^2[T]/\sigma^2[X] + \sigma^2[E]/\sigma^2[X] \text{-----} 2.10$$

### 2.5.1 Reliability

Reliability can be defined as consistency over parallel measures [Crocker & Algina, 1986]. Measures tend to be consistent to the extent that they are free from variance due to random error influences. Reliability can thus be defined as the extent to which a set of measures is free from random error variance and consequently, given Equation 2.10, the ratio of true score variance to observed test score variance. Nunnally [1978, p. 191] for example states:

Random errors of measurement are never completely eliminated; but to portray nature in its ultimate lawfulness, efforts are made to reduce errors as much as possible. To the extent to which measurement error is slight, a measure is set to be *reliable*.

Guion [1965, p.30] in addition states:

Traditionally reliability has been defined as consistency of measurement. It has been virtually a synonym for repeatability. Defined in this way, the concept is certainly useful, but not easily applied to day-to-day measurement problems. The source of the consistency must be considered. ... If the measures are consistent they tend to be free from variance due to random errors. ... As a basic concept, then reliability can be defined as *the extent to which a set of measures is free from random-error variance*.

If a set of measures would be perfectly reliable, the observed test scores would be free of random measurement error and consequently  $\sigma^2[E]$  and thus  $\sigma^2[E]/\sigma^2[X]$  would equal zero and  $\sigma^2[T]/\sigma^2[X]$  would equal unity. Conversely, if a set of measures would be totally unreliable, the observed test score would constitute nothing but random measurement error and consequently  $\sigma^2[T]$  and thus  $\sigma^2[T]/\sigma^2[X]$  would equal zero and  $\sigma^2[E]/\sigma^2[X]$  would equal unity. Consequently, if the ratio of the true score variance to the observed score variance [or conversely, the ratio of the error score variance to the observed score variance] could be estimated, an appropriate indicator/measure of the reliability of a set of score would be available. The quest for a quantitative measure of the reliability of a set of scores can



be approached from the two slightly dissimilar perspectives on inter-individual and intra-individual variability [i.e. by either focussing on the columns or the rows in Table 2.1]. In terms of the inter-individual approach, an estimate of the extent to which the observed measures contain random measurement error [i.e. are unreliable] is obtained through the consistency with which testees maintain their rank-ordered position in the total group on parallel measures [Feldt & Brennan, 1989; Guion, 1965; Stanley, 1971]. In terms of the latter approach, however, an estimate of the extent to which the observed measures contain random measurement error [i.e. are unreliable] is obtained through the fluctuations in observed scores on parallel measures, expressed in score-scale units [Feldt & Brennan, 1989; Stanley, 1971].

Approaching the quest for a quantitative measure of the reliability of a set of scores from an inter-individual perspective on score variability, the covariance and correlation between observed scores and true scores are subsequently examined.

$$\begin{aligned}
 \sigma[X_{ij}, T_{ij}] &= E[X_{ij}T_{ij}] - E[X_{ij}]E[T_{ij}]; i=1, 2, \dots, n; j \\
 &= E[(T_{ij} + E_{ij})T_{ij}] - E[(T_{ij} + E_{ij})]E[T_{ij}] \\
 &= E[T_{ij}^2] + E[E_{ij}T_{ij}] - E[T_{ij}]^2 - E[E_{ij}]E[T_{ij}] \\
 &= \sigma^2[T] + E[E_{ij}T_{ij}] - E[E_{ij}]E[T_{ij}] \text{-----} 2.11.1
 \end{aligned}$$

However, according to Equation 2.2,  $E[E_{ij}] = 0$  and,  $\sigma[T, E] = \rho[T, E]\sigma[T]\sigma[E] = 0$

Since, according to Equation 2.3,  $\rho[T, E] = 0$ .

Hence:

$$\sigma[X_{ij}, T_{ij}] = \sigma[X, T] = \sigma^2[T] \text{-----} 2.11.2$$

However, by definition:

$$\rho^2[X, T] = \sigma^2[X, T] / (\sigma^2[X]\sigma^2[T]) \text{-----} 2.11.3$$

Thus, by substituting in accordance with Equation 2.11.2, Equation 2.11.3 becomes:

$$\begin{aligned}
 \rho^2[X, T] &= \sigma^4[T] / (\sigma^2[X]\sigma^2[T]) \\
 &= \sigma^2[T] / \sigma^2[X] \text{-----} 2.11
 \end{aligned}$$

The square of the correlation between observed scores and true scores equals the ratio of the true score variance to the observed score variance. The quantity,  $\rho[X, T] = \sigma[T] / \sigma[X]$ , is referred to as the index of reliability [Guion, 1965; Gulliksen, 1950; Lord & Novick, 1968]. From an inter-individual variability

perspective, the square of the index of reliability [i.e. interpreted in terms of the coefficient of determination] thus provides a quantitative description of the reliability of a set of scores. From Equation 2.9 and Equation 2.11, Equation 2.12 is obtained.

According to Equation 2.11,

$$\begin{aligned}\rho^2[X,T] &= \sigma^2[T]/\sigma^2[X] \\ &= 1 - (\sigma^2[E]/\sigma^2[X])\end{aligned}\quad 2.12$$

Since according to Equation 2.9,  $\sigma^2[X] = \sigma^2[T] + \sigma^2[E]$

Thus the square of the correlation between the observed scores and true scores could also be interpreted as the proportion of observed test variance remaining once the random error variance has been removed. Using a procedure similar to that followed to obtain Equation 2.11, it could be shown [Equation 2.13] that the square of the correlation between the observed scores and error scores equals the ratio of random error variance to observed score variance.

$$\begin{aligned}\sigma[X_{ij}, E_{ij}] &= E[X_{ij}E_{ij}] - E[X_{ij}]E[E_{ij}]; i=1, 2, \dots, n; j \\ &= E[(T_{ij} + E_{ij})E_{ij}] - E[(T_{ij} + E_{ij})]E[E_{ij}] \\ &= E[E_{ij}^2] + E[E_{ij}T_{ij}] - E[E_{ij}]^2 - E[E_{ij}]E[T_{ij}] \\ &= \sigma^2_E + E[E_{ij}T_{ij}] - E[E_{ij}]E[T_{ij}]\end{aligned}\quad 2.13.1$$

However according to Equation 2.2,  $E[E_{ij}] = 0$  and,  $\sigma[T,E] = \rho[T,E]\sigma[T]\sigma[E] = 0$

Since according to Equation 2.3,  $\rho[T,E] = 0$ .

Hence:

$$\sigma[X_{ij}, E_{ij}] = \sigma[X,E] = \sigma^2[E]\quad 2.13.2$$

However, by definition:

$$\rho^2[X,E] = \sigma^2[X,E]/(\sigma^2[X]\sigma^2[E])\quad 2.13.3$$

Thus, by substituting in accordance with Equation 2.13.2, Equation 2.13.3 becomes:

$$\begin{aligned}\rho^2[X,E] &= \sigma^4[E]/(\sigma^2[X]\sigma^2[E]) \\ &= \sigma^2[E]/\sigma^2[X]\end{aligned}\quad 2.13$$

Although interlocking perfectly with the conceptual definition of reliability, neither the index of reliability [Equation 2.11], nor Equation 2.13, could, however, be computed directly. A practical measure/indicator of the reliability of a set of observed measures can, however, be obtained by expressing the unobservable index of reliability in terms of a parameter of the bivariate observed score distribution.

### 2.5.1.1 Reliability Estimates

Assume two classically parallel measures  $p$  and  $q$ . From the definition of parallel measures follows that  $\mu_p = E[T_{ip}] = \mu_q = E[T_{iq}]$ . According to Equation 2.14 the correlation between two truly parallel measures, applied such that  $T_{ip} = T_{iq}$  over all values of  $i$  [i.e. no practice, fatigue, memory or any other systematic method factor affects the repeated measurement], equals the ratio of the true score variance to the observed score variance.

From the basic definition formula, it follows that:

$$\rho[X_{ip}, X_{iq}] = \sigma[X_{ip}, X_{iq}] / (\sigma[X_{ip}] \sigma[X_{iq}]); i=1, 2, \dots, n; j=p, q \text{-----} 2.14.1$$

However:

$$\begin{aligned} \sigma[X_{ip}, X_{iq}] &= E[X_{ip}X_{iq}] - E[X_{ip}]E[X_{iq}] \\ &= E[(T_{ip} + E_{ip})(T_{iq} + E_{iq})] - E[(T_{ip} + E_{ip})]E[(T_{iq} + E_{iq})] \\ &= E[(T_i + E_{ip})(T_i + E_{iq})] - E[(T_i + E_{ip})]E[(T_i + E_{iq})] \text{-----} 2.14.2 \end{aligned}$$

Since from the tenants of the classical measurement model it follows that:

$$T_{ip} = T_{iq}; i=1, 2, \dots, n; j=p, q$$

Thus:

$$\begin{aligned} \sigma[X_{ip}, X_{iq}] &= E[T_i T_i] + 2E[E_{ij} T_i] + E[E_{ip} E_{iq}] - [E[T_i] + E[E_{ip}]] [E[T_i] + E[E_{iq}]] \\ &= E[T_i^2] + 2E[E_{ij} T_i] + E[E_{ip} E_{iq}] - E[T_i]E[T_i] - E[E_{ip}]E[T_i] - E[T_i]E[E_{iq}] - \\ &\quad E[E_{iq}]E[E_{ip}] \text{-----} 2.14.3 \end{aligned}$$

However according to Equation 2.2,  $E[E_{ip}] = E[E_{iq}] = 0$

and, according to Equations 2.3 and 2.4,  $E[E_{ij}T_i] = E[E_{ip}E_{iq}] = 0$

Thus Equation 2.14.2 can be rewritten as:

$$\begin{aligned}\sigma[X_{ip}, X_{iq}] &= E[T_i^2] - E[T_i]E[T_i] \\ &= E[T_i^2] - (E[T_i])^2 \\ &= \sigma^2[T] \text{-----} 2.14.4\end{aligned}$$

Thus Equation 2.14.1 can be rewritten as:

$$\rho[X_{ip}, X_{iq}] = \sigma^2[T] / (\sigma[X_{ip}]\sigma[X_{iq}]) \text{-----} 2.14.5$$

However,  $\sigma[X_{ip}] = \sigma[X_{iq}] = \sigma[X]$

Thus:

$$\rho[X_{ip}, X_{iq}] = \sigma^2[T] / \sigma^2[X] \text{-----} 2.14$$

Equation 2.14 is often referred to as the coefficient of precision [Coombs, 1950; Lord & Novick, 1968]. The coefficient of precision could thus be defined as the correlation between two truly classically parallel measurements [ i.e.  $T_{ip} = T_{iq}$  and  $\sigma^2[E_{ip}] = \sigma^2[E_{iq}]$ ] obtained with infinitesimal elapsed time between measurements [Coombs, 1950; Crocker & Algina, 1986; Lord & Novick, 1968; Novick, 1966;]. Combining Equations 2.13 and 2.14 results in the deduction of an estimate of the index of reliability through the square root of the correlation between the observed scores obtained over n testees from two parallel measures as shown by Equation 2.15.

According to Equation 2.13,  $\rho^2[X, T] = \sigma^2[T] / \sigma^2[X]$

According to Equation 2.14,  $\rho[X_{ip}, X_{iq}] = \sigma^2[T] / \sigma^2[X]; i=1, 2, \dots n; j=p, q$

Write:

$$\rho[X_{ip}, X_{iq}] \equiv \rho_{ttX}$$

Thus:

$$\rho_{ttX} = \rho^2[X, T] \text{-----} 2.15$$

Contingent on the assumption that at least one pair of classically parallel measures are available, Equation 2.15 equates the conceptually pleasing, but empirically unobservable, quantity  $\rho^2[X, T]$  to the observable quality  $\rho_{ttX}$  [Lord & Novick, 1968].

Rearranging the terms in Equation 2.15 through cross multiplication, Equation 2.16 is obtained.

Write  $\rho[X_{ip}, X_{iq}] \equiv \rho_{ttX}; i=1, 2, \dots, n; j=p, q$

Thus, according to Equation 2.14,  $\rho_{ttX} = \sigma^2[T]/\sigma^2[X]$

Thus:

$$\sigma^2[T] = \rho_{ttX}\sigma^2[X] \text{-----} 2.16$$

According to Equation 2.16 the true score variance is equal to, and thus can be estimated by, the product of the observed score variance and the correlation between parallel measurements. The error score variance, in contrast, is equal to the product of the observed score variance and one minus the correlation between parallel measurements, as shown by Equation 2.17.

According to Equation 2.16,  $\sigma^2[T] = \rho_{ttX}\sigma^2[X]$

According to Equation 2.9,  $\sigma^2[X] = \sigma^2[T] + \sigma^2[E]$

Thus, by substituting Equation 2.16 into Equation 2.9 it follows that::

$$\sigma^2[X] = \rho_{ttX}\sigma^2[X] + \sigma^2[E] \text{-----} 2.17.1$$

$$\begin{aligned} \text{Thus } \sigma^2[E] &= \sigma^2[X] - \rho_{ttX}\sigma^2[X] \\ &= \sigma^2[X][1 - \rho_{ttX}] \text{-----} 2.17 \end{aligned}$$

By taking square roots, Equation 2.17 can be written as Equation 2.18, which defines the standard error of measurement.

According to Equation 2.17,  $\sigma^2[E] = \sigma^2[X] - \rho_{ttX}\sigma^2[X]$

By taking square roots it thus follows that::

$$\sigma[E] = \sigma[X]\sqrt{1 - \rho_{ttX}} \text{-----} 2.18$$



The standard error of measurement and the error variance [Equation 2.17] provide quantitative descriptions of the reliability of a set of observed scores from an intra-individual perspective. Both measures provide quantitative descriptions, expressed in scale-score units, of the average dispersion in the observed scores over  $k$  parallel test forms around the true scores. Homogeneity of error variance across the full range of true score levels [i.e. across testees] could be an additional assumption of the classical measurement model that would, to the extent that the homoscedasticity assumption is in fact accurate, permit the derivation of convenient additional results [Feldt & Brennan, 1989]. There, however, exists considerable evidence that departures from homoscedasticity are not uncommon in psychological measurement [Feldt & Brennan, 1989]. Specifically, the error variance associated with the lower and higher true scores tend to be smaller than that associated with scores in the middle of the true score range [Feldt & Brennan, 1989]. Although a constant for testee  $i$ , the error variance and standard error of measurement for fixed testee  $i$  [Lord & Novick, 1968] represent variables when taken over  $i$ , thus necessitating the definition of these measures with respect to the measuring instrument in terms of their expected value. Equation 2.17 and Equation 2.18 return the expected value of the error variance and standard error of measurement taken over testees. These equations, however, also apply if the homoscedasticity assumption would in fact be valid [Feldt & Brennan, 1989].

From an inter-individual perspective, three practical methods for the estimation of the reliability of a set of observed scores with parallel measurements are available [Lord & Novick, 1968]:

- the test-retest method;
- the internal analysis method; and
- the parallel forms method.

Different conceptions of parallel measurements do, however, exist. These different conceptions of the parallel measurements constitute, in conjunction with the aforementioned array of methods, a logical framework for organising different empirical reliability estimates [Feldt & Brennan, 1989]. The following interpretations of parallel measurement could be distinguished, ordered from the most to the least demanding and stringent.

- classically parallel measurements;
- tau-equivalent measurements;
- essentially tau-equivalent measurements;
- congeneric measurements; and
- multi-factor congeneric measurements.

The rather stringent, and simultaneously somewhat unrealistic, requirements set forth by the classical measurement model regarding parallel measurements were outlined above. Classically parallel measurements are characterised by identical  $\mu[X_j]$ ,  $\sigma^2[X_j]$ ,  $\sigma^2[T_j]$ ,  $\sigma^2[E_j]$  and  $\rho[X_p, X_q]$  for all  $j = 1, \dots, p, q, \dots, k$  and, for every testee  $i$ , a constant true score  $T_{ij}$  over all values of  $j$ . By tau equivalent

measurements is meant measurements with equal true scores but [possibly] different error variances and thus [possibly] different observed score variances [Feldt & Brennan, 1989; Gilmer & Feldt, 1983]. According to Lord and Novick [1968] distinct measurements  $X_{ip}$  and  $X_{iq}$  may be considered  $\tau$ -equivalent if, for all  $i$  and  $j$ ,  $\tau_{ip} = \tau_{iq}$ , and thus,  $E[X_{ip}] = E[X_{iq}]$ . In contrast, distinct measurements  $X_{ip}$  and  $X_{iq}$  may be considered essentially tau-equivalent if, for all  $i$  and  $j$ ,  $\tau_{ip} = a + \tau_{iq}$  [Lord & Novick, 1968; Zimmerman, Zumbo & Lalonde, 1993]. Essentially  $\tau$ -equivalent measurements thus are characterised by true scores that are permitted to vary across measurements while correlating perfectly across measurements. Furthermore, the  $\tau$ -equivalence requirement that, for all  $i$  and  $j$ ,  $E[X_{ip}] = E[X_{iq}]$ , also no longer applies. Congeneric measurements represent an additional easing off on the prerequisites posed for parallel measurements. Congeneric measurements permit more extreme true scores variation across measurements than essentially  $\tau$ -equivalent measurements, by requiring, for all  $i$  and  $j$ , that  $\tau_{ip} = a + b\tau_{iq}$ . The coefficients  $a$  and  $b$  are constant over  $i$  but permitted to vary over  $j$ . The coefficient  $b$  represents the effect of increasing or decreasing the number of components composite measurements are comprised of. True scores, although varying across measurements, however, still correlate perfectly across measurements. Observed score variance will reflect inequalities in both  $\sigma^2[T]$  and  $\sigma^2[E]$ . Multi-factor congeneric measurements conceive the true score component of a measurement as a weighted sum of  $g$  systematic components  $T_1$  through  $T_g$ . The weights of the true score components are permitted to vary over components and measurements. Thus different combinations of the same  $g$  systematic true score components bring forth the true score on repeated measurements. Consequently the true scores on multi-factor congeneric measurements do not correlate perfectly.

#### 2.5.1.1.1 Test-Retest Method

The test-retest method involves administering the same test to a sample of testees twice. The resultant observations on the first and second application are then correlated. The correlation may be taken, with some caution, as an approximation of coefficient of precision [Lord & Novick, 1968]. However, contrary to the coefficient of precision, transfer effects [practice, memory, fatigue, boredom] tend to affect the second administration [Ghiselli, Campbell & Zedeck, 1978] and thus the correlation between repeated measurements. The time span between repeated administrations also plays a significant role, firstly because it correlates negatively with the impact of the transfer effects and secondly, because it affects the probability of changes in the relevant and/or non-relevant factors affecting test behaviour. The problem with the test-retest method of reliability estimation thus lies therein that the resultant coefficient of stability reflects changes in testees over time and the lack of parallelism in the repeated measurements. The coefficient of precision reflects the correlation between truly parallel measurements obtained without changes in the testees true score. It is, however, difficult to predict

whether, on balance, these effects would tend to increase or decrease the correlation between the two measurements [Kuder & Richardson, 1937; Lord & Novick, 1968]. Furthermore, purely from a practical administrative perspective, recapture of testees for a repetition of measurement often presented a formidable obstacle.

#### 2.5.1.1.2 Internal Analysis Method

The second practical method for the estimation of the reliability of a set of observed scores with parallel measurements requires an internal analysis of the variances and covariances of parallel components comprising the test. This method thus requires only a single administration of the test [Crocker & Algina, 1986]. One way of creating parallel components is to divide the items comprising the test into two parts [i.e. split halves]. Assume therefore two parallel composite measurements,  $X_1$  and  $X_2$ , each composed of two parallel measurements [ $X_{11}$  and  $X_{12}$ , and,  $X_{21}$  and  $X_{22}$ ]. Assume each composite measurement to be additively determined from the two component parts, such that  $X_1 = X_{11} + X_{12}$  and  $X_2 = X_{21} + X_{22}$ .

The correlation between the two parallel composite measures can be written as:

$$\begin{aligned}\rho[X_1, X_2] &= \rho[(X_{11} + X_{12}), (X_{21} + X_{22})] \\ &= \sigma[(X_{11} + X_{12}), (X_{21} + X_{22})] / \sigma[(X_{11} + X_{12})] \sigma[(X_{21} + X_{22})] \\ &= \sigma[(X_{11} + X_{12}), (X_{21} + X_{22})] / \sqrt{\{\sigma[(X_{11} + X_{12})]\}^2 \{\sigma[(X_{21} + X_{22})]\}^2} \text{-----} 2.19.1\end{aligned}$$

The covariance term can be written as:

$$\begin{aligned}\sigma[(X_{11} + X_{12}), (X_{21} + X_{22})] &= \sigma[X_{11}, X_{21}] + \sigma[X_{11}, X_{22}] + \sigma[X_{12}, X_{21}] + \sigma[X_{12}, X_{22}] \\ &= 4\sigma^2[X_{11}]r[X_{11}, X_{12}] \\ &= 2\sigma^2[X_{11}](2r[X_{11}, X_{12}]) \text{-----} 2.19.2\end{aligned}$$

$$\begin{aligned}\text{Since } \sigma[X_{11}, X_{12}] &= \rho[X_{11}, X_{12}]\sigma[X_{11}]\sigma[X_{12}] \\ &= \rho[X_{11}, X_{12}]\sigma^2[X_{11}] \text{ for all covariance terms}\end{aligned}$$

and since the observed score variance and inter-correlations of all parallel measures are equal

$$\begin{aligned}&\sqrt{\{\sigma[(X_{11} + X_{12})]\}^2 \{\sigma[(X_{21} + X_{22})]\}^2} \\ &= \sqrt{\{\sigma^2[X_{11}] + \sigma^2[X_{12}] + 2\sigma[X_{11}, X_{12}]\} \{\sigma^2[X_{21}] + \sigma^2[X_{22}] + 2\sigma[X_{21}, X_{22}]\}} \\ &= \sqrt{\{(2\sigma^2[X_{11}] + 2\rho[X_{11}, X_{12}]\sigma^2[X_{11}]) (2\sigma^2[X_{11}] + 2\rho[X_{11}, X_{12}]\sigma^2[X_{11}])\}}\end{aligned}$$

$$\begin{aligned}
 &= 2\sigma^2[X_{11}] + 2\rho[X_{11}, X_{12}] \sigma^2[X_{11}] \\
 &= 2\sigma^2[X_{11}](1 + \rho[X_{11}, X_{12}]) \text{-----} 2.19.3
 \end{aligned}$$

$$\begin{aligned}
 \text{Since } \sigma[X_{11}, X_{12}] &= \rho[X_{11}, X_{12}] \sigma[X_{11}] \sigma[X_{12}] \\
 &= \rho[X_{11}, X_{12}] \sigma^2[X_{11}] \text{ for all covariance terms}
 \end{aligned}$$

and since the observed score variance and inter-correlations of all parallel measures are equal

Thus:

$$\begin{aligned}
 \rho[(X_{11}+X_{12}), (X_{21}+X_{22})] &= 2\sigma^2[X_{11}](2\rho[X_{11}, X_{12}]) / 2\sigma^2[X_{11}](1+\rho[X_{11}, X_{12}]) \\
 &= 2\rho[X_{11}, X_{12}] / (1+\rho[X_{11}, X_{12}]) \text{-----} 2.19.4
 \end{aligned}$$

Write:

$$\begin{aligned}
 \rho_{ttX} &= \rho[X_1, X_2] \\
 &= 2\rho[X_{11}, X_{12}] / (1+\rho[X_{11}, X_{12}]) \text{-----} 2.19
 \end{aligned}$$

Equation 2.19 returns the reliability coefficient of a composite test comprising two parallel components from the reliability coefficient applicable to the two parallel components derived from a single application of the composite. Equation 2.19 is generally known as the Spearman-Brown prediction/prophesy formula. The Spearman-Brown formula is based on the assumption of strictly parallel measures. To the extent that the split halves are in fact not truly parallel, the Spearman-Brown formula will underestimate the coefficient of precision [Lord & Novick, 1968]. The process of dividing the test into component parts thus should proceed with circumspection. Random assignment [with or without matching on item statistics] or the assignment of items with the odd-even method generally tend to produce approximately parallel split halves. The practice of dividing the test into split halves by placing the first  $k/2$  items in the first component and the second  $k/2$  items in the second component should, however, be avoided. Due to the common practice of placing test items in order of increasing difficulty, such a procedure would clearly result in nonparallel splits.

A deficiency of the split-half procedure, irrespective of the specific reliability coefficient calculated, is that no unique estimate of the reliability coefficient is obtained since there are  $\frac{1}{2}k![(\frac{1}{2}k)!]^2$  different ways of dividing a  $k$ -item test into split halves [Crocker & Algina, 1986]. Generalising the Spearman-Brown formula to a procedure where the test is divided in  $k$  parallel components would theoretically present one possible solution. Equation 2.20 derives the generalised Spearman-Brown formula. Assume therefore  $k$  parallel item measurements,  $X_1, X_2, \dots, X_k$  with true scores  $T_1, T_2, \dots, T_k$ . Assume

the composite total test measurement to be additively determined from the  $k$  parallel component parts, such that  $X = [X_1 + X_2 + \dots + X_k] = \Sigma X_i$  and  $T = \Sigma T_i$ .

According to Equations 2.11 and 2.14:

$$\begin{aligned}\rho^2[X,T] &= \sigma^2[T]/\sigma^2[X] \\ &= \rho[X_{ip},X_{iq}]\end{aligned}$$

Write:

$$\rho[X_{ip},X_{iq}] \equiv \rho_{ttX}$$

$$\text{Thus } \rho_{ttX} = \sigma^2[T]/\sigma^2[X]$$

Because of the assumption that the composite total test measurement to be additively determined from the  $k$  parallel component parts, such that  $X = [X_1 + X_2 + \dots + X_k] = \Sigma X_i$ , the test variance can be analysed as follows:

$$\begin{aligned}\sigma^2[X] &= \sigma^2[\Sigma X_i] \\ &= \Sigma \sigma^2[X_i] + \Sigma \Sigma \sigma[X_i, X_j]; i \neq j \\ &= \Sigma \sigma^2[X_i] + \Sigma \Sigma \rho[X_i, X_j] \sigma[X_i] \sigma[X_j]; i \neq j\end{aligned}\quad \text{----- 2.20.1}$$

$$\text{Since } \rho[X_i, X_j] = \sigma[X_i, X_j] / (\sigma[X_i] \sigma[X_j])$$

Thus:

$$\sigma^2[X] = k\sigma^2[X_i] + k[k-1]\rho[X_i, X_j]\sigma^2[X_i]; i \neq j\quad \text{----- 2.20.2}$$

Since all components constitute parallel measurements all  $\rho[X_i, X_j]$  are equal and  $\sigma[X_i] = \sigma[X_j]$  for all  $i$  and  $j$

Simplifying Equation 2.20.2 results in:

$$\sigma^2[X] = k\sigma^2[X_i](1 + [k-1]\rho[X_i, X_j])\quad \text{----- 2.20.3}$$

Because of the assumption that the composite total test measurement to be additively determined from the  $k$  parallel component parts, such that  $T = \Sigma T_i$ , the true score variance can be analysed as follows:

$$\sigma^2[T] = \sigma^2[\Sigma T_i]$$



$$\begin{aligned}
&= \Sigma \sigma^2[T_i] + \Sigma \Sigma \sigma[T_i, T_j]; i \neq j \\
&= k\sigma^2[T_i] + k[k-1]\sigma^2[T_i] \text{-----} 2.20.4
\end{aligned}$$

Since, all components constitute parallel measurements and thus  $\sigma[T_i, T_j] = \sigma^2[T_i]$  for all  $i$  and  $j$

Thus, by multiplying through the last term of Equation 2.20.4:

$$\begin{aligned}
\sigma^2[T] &= k\sigma^2[T_i] + k^2\sigma^2[T_i] - k\sigma^2[T_i] \\
&= k^2\sigma^2[T_i] \text{-----} 2.20.5
\end{aligned}$$

Thus:

$$\begin{aligned}
\rho_{ttX} &= \sigma^2[T]/\sigma^2[X] \\
&= k^2\sigma^2[T_i]/(k\sigma^2[X_i]\{1 + [k-1]\rho[X_i, X_j]\}) \\
&= k\sigma^2[T_i]/(\sigma^2[X_i]\{1 + [k-1]\rho[X_i, X_j]\}) \\
&= k\rho[X_i, X_j]/(1 + [k-1]\rho[X_i, X_j]) \text{-----} 2.20
\end{aligned}$$

The generalised Spearman-Brown formula [Equation 2.20] still assumes strictly parallel components. The probability of  $k$  test items satisfying this assumption seems, however, rather remote. Additional procedures were, consequently, subsequently proposed to overcome the deficiency associated with the split-halves procedure [Cronbach, 1951; Feldt, 1975; Guttman, 1945; Hoyt, 1941; Kuder & Richardson, 1937; Raju, 1977; Rulon, 1939].

Let  $X_1$  and  $X_2$  be essentially tau-equivalent rather than classically parallel component measurements with true scores  $T_1$  and  $T_2$ . Let  $X = X_1 + X_2$  be a composite measurement with true score  $T = \Sigma T_i$ .

A squared non-zero difference necessarily must be positive.

Consequently the following inequality can be stated:

$$(\sigma[T_1] - \sigma[T_2])^2 \geq 0 \text{-----} 2.21.1$$

Thus:

$$\sigma^2[T_1] + \sigma^2[T_2] - 2\sigma[T_1]\sigma[T_2] \geq 0 \text{-----} 2.21.2$$

Thus:

$$\sigma^2[T_1] + \sigma^2[T_2] \geq 2\sigma[T_1]\sigma[T_2] \text{-----} 2.21.3$$

However:

$$\rho[T_1, T_2] = |\sigma[T_1, T_2]| / (\sigma[T_1]\sigma[T_2]) \leq 1 \text{-----} 2.21.4$$

Thus it follows that 2.21.4 can only be true if:

$$\begin{aligned} \sigma[T_1]\sigma[T_2] &\geq |\sigma[T_1, T_2]| \\ &\geq \sigma[T_1, T_2] \text{-----} 2.21.5 \end{aligned}$$

Thus:

$$2\sigma[T_1]\sigma[T_2] \geq 2\sigma[T_1, T_2] \text{-----} 2.21.6$$

By combining Equations 2.21.3 and 2.21.6 it follows that:

$$\sigma^2[T_1] + \sigma^2[T_2] \geq 2\sigma[T_1, T_2] \text{-----} 2.21.7$$

By adding  $2\sigma[T_1, T_2]$  on both sides of the equality:

$$\begin{aligned} \sigma^2[T_1] + \sigma^2[T_2] + 2\sigma[T_1, T_2] &\geq 2\sigma[T_1, T_2] + 2\sigma[T_1, T_2] \\ &\geq 4\sigma[T_1, T_2] \text{-----} 2.21.8 \end{aligned}$$

Because of the assumption that  $X = X_1 + X_2$  is a composite measurement with true score  $T = \Sigma T_i$ , The true score variance can be analysed as follows:

$$\begin{aligned} \sigma^2[T] &= \sigma^2[T_1 + T_2] \\ &= \sigma^2[T_1] + \sigma^2[T_2] + 2\sigma[T_1, T_2] \text{-----} 2.21.9 \end{aligned}$$

Combining Equations 2.21.8 and 2.21.9 consequently results in:

$$\sigma^2[T] \geq 4\sigma[T_1, T_2] \text{-----} 2.21.10$$

Dividing both sides of Equation 2.21.10 with  $\sigma^2[X]$  results in:

$$\begin{aligned} \sigma^2[T] / \sigma^2[X] &\geq 4\sigma[T_1, T_2] / \sigma^2[X] \\ &\geq 2(2\sigma[T_1, T_2] / \sigma^2[X]) \text{-----} 2.21.11 \end{aligned}$$

However:

$$\begin{aligned}\sigma[X_1, X_2] &= \sigma[(T_1 + E_1), (T_2 + E_2)] \\ &= \sigma[T_1, T_2] + \sigma[T_1, E_2] + \sigma[E_1, T_2] + \sigma[E_1, E_2] \\ &= \sigma[T_1, T_2] \text{-----} 2.21.12\end{aligned}$$

Since, Equations 2.3 and 2.4 would imply that  $\sigma[T_1, E_2] = \sigma[E_1, T_2] = \sigma[E_1, E_2] = 0$

Combining Equations 2.21.11 and 2.21.12 therefore results in:

$$\sigma^2[T]/\sigma^2[X] \geq 2(2\sigma[X_1, X_2]/\sigma^2[X]) \text{-----} 2.21.13$$

Because of the assumption that  $X_1$  and  $X_2$  are essentially tau-equivalent component measurements such that  $X = X_1 + X_2$  the test variance can be analysed as follows:

$$\begin{aligned}\sigma^2[X] &= \sigma^2[X_1 + X_2] \\ &= \sigma^2[X_1] + \sigma^2[X_2] + 2\sigma[X_1, X_2] \text{-----} 2.21.14\end{aligned}$$

Isolating the covariance term in Equation 2.21.14 results in:

$$2\sigma[X_1, X_2] = \sigma^2[X] - \sigma^2[X_1] - \sigma^2[X_2] \text{-----} 2.21.15$$

Substituting Equation 2.21.15 in Equation 2.21.13 results in:

$$\sigma^2[T]/\sigma^2[X] \geq 2\{(\sigma^2[X] - \sigma^2[X_1] - \sigma^2[X_2])/\sigma^2[X]\} \text{-----} 2.21.16$$

Therefore:

$$\begin{aligned}\rho^2[X, T] &= \rho_{ttX} \\ &= \sigma^2[T]/\sigma^2[X] \\ &\geq 2\{1 - (\sigma^2[X_1] + \sigma^2[X_2])/\sigma^2[X]\} \text{-----} 2.21\end{aligned}$$

Equation 2.21 is generally known as coefficient  $\alpha$  [Cronbach, 1951] as it would apply to a two component composite. Coefficient  $\alpha$  will return a lower bound of the coefficient of precision unless the two components may be considered essentially tau-equivalent, in which case Equation 2.21 holds as an equality. Equation 2.21 is also known as the  $\lambda_4$  Guttman split-halves reliability coefficient [Guttman, 1945]. Equation 2.21 may be expressed in two alternative, but equivalent, algebraic forms.

The formula, attributed to Flanagan [Rulon, 1939], is indicated by the argument used to derive the Cronbach/Guttman equation and shown as Equation 2.22.

From Equation 2.21, it follows that:

$$\begin{aligned}\rho^2[X,T] &= \sigma^2[T]/\sigma^2[X] \\ &\geq 2(2\sigma[X_1,X_2]/\sigma^2[X]) \\ &\geq 4\sigma[X_1,X_2]/\sigma^2[X] \text{-----} 2.22\end{aligned}$$

The formula, attributed to Rulon [Rulon, 1939], is shown as Equation 2.23. Let  $X_1$  and  $X_2$  be essentially  $\tau$ -equivalent component measurements with true scores  $T_1$  and  $T_2$ . Let  $X = X_1 + X_2$  be a composite measurement with true score  $T = \Sigma T_i$ .

Using the Flanagan expression [Equation 2.22] as point of departure:

$$\begin{aligned}\rho^2[X,T] &= \sigma^2[T]/\sigma^2[X] \\ &\geq 4\sigma[X_1,X_2]/\sigma^2[X] \\ &\geq (2\sigma[X_1,X_2] + 2\sigma[X_1,X_2])/\sigma^2[X] \text{-----} 2.23.1\end{aligned}$$

$$\text{Since } 4\sigma[X_1,X_2] = 2\sigma[X_1,X_2] + 2\sigma[X_1,X_2]$$

Because of the assumption that  $X_1$  and  $X_2$  are essentially tau-equivalent component measurements such that  $X = X_1 + X_2$  the test variance can be analysed as was done earlier in Equation 2.21.14:

$$\begin{aligned}\sigma^2[X] &= \sigma^2[X_1 + X_2] \\ &= \sigma^2[X_1] + \sigma^2[X_2] + 2\sigma[X_1,X_2] \text{-----} 2.23.2\end{aligned}$$

Isolating the covariance term as in Equation 2.21.15, again results in:

$$2\sigma[X_1,X_2] = \sigma^2[X] - \sigma^2[X_1] - \sigma^2[X_2] \text{-----} 2.23.3$$

Substituting one covariance term in Equation 2.23.1 with Equation 2.23.3:

$$\begin{aligned}\rho^2[X,T] &\geq (\sigma^2[X] - \sigma^2[X_1] - \sigma^2[X_2] + 2\sigma[X_1,X_2])/\sigma^2[X] \\ &\geq 1 - (\sigma^2[X_1] + \sigma^2[X_2] - 2\sigma[X_1,X_2])/\sigma^2[X] \text{-----} 2.23.4\end{aligned}$$

$$\text{Let } D = X_1 - X_2$$

$$\begin{aligned}\text{Then, } \sigma^2[D] &= \sigma^2[X_1 - X_2] \\ &= \sigma^2[X_1] + \sigma^2[X_2] - 2\sigma[X_1, X_2] \text{-----} 2.23.5\end{aligned}$$

Substituting Equation 2.23.5 in Equation 2.23.4 results in:

$$\begin{aligned}\rho^2[X, T] &\geq 1 - (\sigma^2[X_1] + \sigma^2[X_2] - 2\sigma[X_1, X_2]) / \sigma^2[X] \\ &\geq 1 - (\sigma^2[D] / \sigma^2[X]) \\ &\geq 1 - (\sigma^2[X_1 - X_2] / \sigma^2[X]) \text{-----} 2.23\end{aligned}$$

If the assumption could further be made that the two components  $X_1$  and  $X_2$  are classically parallel measurements, the Rulon reliability formula [Equation 2.23] could be derived as an equality as shown by Equation 2.24.

Because of the assumption that  $X_1$  and  $X_2$  are classically parallel component measurements such that  $X = X_1 + X_2$  the test variance can be analysed as follows:

$$\begin{aligned}\sigma^2[X] &= \sigma^2[X_1 + X_2] \\ &= \sigma^2[X_1] + \sigma^2[X_2] + 2\sigma[X_1, X_2] \\ &= \sigma^2[X_1] + \sigma^2[X_2] + 2\rho[X_1, X_2]\sigma[X_1]\sigma[X_2] \\ &= 2\sigma^2[X_1] + 2\rho[X_1, X_2]\sigma^2[X_1] \\ &= 2\sigma^2[X_1](1 + \rho[X_1, X_2]) \text{-----} 2.24.1\end{aligned}$$

$$\begin{aligned}\text{Since } \rho[X_1, X_2] &= \sigma[X_1, X_2] / (\sigma[X_1]\sigma[X_2]); \text{ and} \\ \sigma^2[X_1] &= \sigma^2[X_2]\end{aligned}$$

$$\text{Equation 2.18 states that } \sigma[E] = \sqrt{(\sigma^2[X](1 - \rho_{ttX}))}$$

Squaring Equation 2.18 and substituting the test variance term with Equation 2.24.1 results in:

$$\begin{aligned}\sigma^2[E] &= \sigma^2[X](1 - \rho_{ttX}) \\ &= 2\sigma^2[X_1](1 + \rho[X_1, X_2])(1 - \rho_{ttX}) \text{-----} 2.24.2\end{aligned}$$

Substituting  $\rho_{ttX}$  in Equation 2.24.2 with Equation 2.19 results in:

$$\begin{aligned}\sigma^2[E] &= 2\sigma^2[X_1](1 + \rho[X_1, X_2])(1 - 2\rho[X_1, X_2]) / (1 + \rho[X_1, X_2]) \\ &= 2\sigma^2[X_1](1 + \rho[X_1, X_2])(1 + \rho[X_1, X_2] - 2\rho[X_1, X_2]) / (1 + \rho[X_1, X_2]) \text{-----} 2.24.3\end{aligned}$$

The term  $(1 + \rho[X_1, X_2])$  cancels out, resulting in:



$$\begin{aligned}\sigma^2[E] &= 2\sigma^2[X_1](1 + \rho[X_1, X_2] - 2\rho[X_1, X_2]) \\ &= 2\sigma^2[X_1](1 - \rho[X_1, X_2])\end{aligned}\quad 2.24.4$$

Let  $D = X_1 - X_2$

The variance of the difference score can be analysed as follows:

$$\begin{aligned}\sigma^2[D] &= \sigma^2[X_1 - X_2] \\ &= \sigma^2[X_1] + \sigma^2[X_2] - 2\rho[X_1, X_2]\sigma[X_1]\sigma[X_2] \\ &= 2\sigma^2[X_1] - 2\rho[X_1, X_2]\sigma^2[X_1] \\ &= 2\sigma^2[X_1](1 - \rho[X_1, X_2])\end{aligned}\quad 2.24.5$$

Comparing Equations 2.24.4 and 2.24.5 results in the conclusion that:

$$\sigma^2[D] = \sigma^2[E]\quad 2.24.6$$

Utilising Equations 2.11 and 2.12 and substituting  $\sigma^2[E]$  with  $\sigma^2[D]$  in Equation 2.12 in accordance with Equation 2.24.6:

$$\begin{aligned}\rho^2[X, T] &= \sigma^2[T]/\sigma^2[X] \\ &= 1 - (\sigma^2[E]/\sigma^2[X]) \\ &= 1 - (\sigma^2[D]/\sigma^2[X])\end{aligned}\quad 2.24.7$$

Substituting  $\sigma^2[D]$  in Equation 2.24.7 with its alternative representation  $\sigma^2[X_1 - X_2]$ , results in:

$$\rho_{TTX} = \sigma^2[T]/\sigma^2[X] = 1 - (\sigma^2[X_1 - X_2]/\sigma^2[X])\quad 2.24$$

Furthermore, given the assumption that the two components  $X_1$  and  $X_2$  are parallel measurements, Equation 2.21 reduces to the Spearman-Brown formula [Equation 2.19], as shown by Equation 2.25.

Using Equation 2.21 as the basic point of departure:

$$\rho^2[X, T] = \sigma^2[T]/\sigma^2[X] \geq 2\{1 - (\sigma^2[X_1] + \sigma^2[X_2])/\sigma^2[X]\}\quad 2.25.1$$

If  $X_1$  and  $X_2$  are parallel measurements then:

$$\sigma^2[X_1] = \sigma^2[X_2]\quad 2.25.2$$

Applying Equation 2.25.2 to Equation 2.25.1 results in:

$$\rho^2[X,T] \geq 2\{1-(2\sigma^2[X_1])/\sigma^2[X]\} \text{-----} 2.25.3$$

Because of the assumption that  $X_1$  and  $X_2$  are classically parallel component measurements such that  $X = X_1 + X_2$  the test variance can be analysed as follows:

$$\begin{aligned} \sigma^2[X] &= \sigma^2[X_1+X_2] \\ &= \sigma^2[X_1]+\sigma^2[X_2]+\sigma[X_1,X_2] \\ &= \sigma^2[X_1]+\sigma^2[X_2]+2\rho[X_1,X_2]\sigma[X_1]\sigma[X_2] \text{-----} 2.25.4 \end{aligned}$$

$$\text{Since, } \rho[X_1,X_2] = \sigma[X_1,X_2]/(\sigma[X_1]\sigma[X_2])$$

Applying Equation 2.25.2 to Equation 2.25.4 results in:

$$\begin{aligned} \sigma^2[X] &= 2\sigma^2[X_1]+2\rho[X_1,X_2]\sigma^2[X_1] \\ &= 2\sigma^2[X_1](1 + \rho[X_1,X_2]) \text{-----} 2.25.5 \end{aligned}$$

Employing Equation 2.25.3 and substituting  $\sigma^2[X]$  with Equation 2.25.5 results in:

$$\begin{aligned} \rho^2[X,T] &\geq 2\{1-(2\sigma^2[X_1])/\sigma^2[X]\} \\ &\geq 2\{1-(2\sigma^2[X_1])/2\sigma^2[X_1](1 + \rho[X_1,X_2])\} \\ &\geq 2\{2\sigma^2[X_1](1 + \rho[X_1,X_2])-(2\sigma^2[X_1])/2\sigma^2[X_1](1 + \rho[X_1,X_2])\} \text{-----} 2.25.6 \end{aligned}$$

By factoring out the term  $2\sigma^2[X_1]$ , Equation 2.25.6 can be simplified as follows:

$$\begin{aligned} \rho^2[X,T] &\geq 2\{2\sigma^2[X_1]\{(1 + \rho[X_1,X_2])-1\}/2\sigma^2[X_1](1 + \rho[X_1,X_2])\} \\ &\geq 2\{((1 + \rho[X_1,X_2])-1)/(1 + \rho[X_1,X_2])\} \\ &\geq 2\{\rho[X_1,X_2]/(1 + \rho[X_1,X_2])\} \\ &\geq 2\rho[X_1,X_2]/(1 + \rho[X_1,X_2]) \text{-----} 2.25.7 \end{aligned}$$

Consequently, Equation 2.21 reduces to the Spearman-Brown formula if the two component  $X_1$  and  $X_2$  are classically parallel measurements:

$$2\{1-(\sigma^2[X_1]+\sigma^2[X_2])/\sigma^2[X]\} = 2\rho[X_1,X_2]/(1 + \rho[X_1,X_2]) \text{-----} 2.25$$

Raju [1970] derived a reliability coefficient for two congeneric measurements of known length, shown as Equation 2.26.

Let  $k_1$  and  $k_2$  represent the lengths of the two components expressed in an appropriate metric such that  $k_1 \neq k_2$  and  $k_1 + k_2 = k$ . Let the items of the total test be denoted by 1, 2, ...,  $i$ ,  $j$ , ...,  $k$ ; the items of the first component/part by 1, 2, ...,  $m$ ,  $n$ , ...,  $k_1$ ; and the items of the second component/part by 1, 2, ...,  $u$ ,  $v$ , ...,  $k_2$ .

Let  $\lambda_j$  represent constants [ $\lambda_j > 0$ ;  $\sum \lambda_j = 1$ ] which reflect the proportions of the total test true score contributed by the 2 components. . Let  $b_j$  [ $\sum b_j = 0$ ] represent constants which provide for differences in component mean scores not attributable to differences in component length.

Taking Equation 2.1 as the basic point of departure:

$$\text{Let } X = T + E \text{-----} 2.26.1$$

$$T = \sum T_i \text{-----} 2.26.2$$

$$E = \sum E_i \text{-----} 2.26.3$$

In terms of the congeneric measurement model the component observed scores can be broken down into the following parts:

$$X_1 = \lambda_1 T + b_1 + E_1 \text{-----} 2.26.4$$

$$X_2 = \lambda_2 T + b_2 + E_2 \text{-----} 2.26.5$$

Based on Equation 2.26.2 the true score variance can be analysed as follows:

$$\begin{aligned} \sigma^2[T] &= \sigma^2[T_1 + T_2 + \dots + T_i + T_j + \dots + T_k] \\ &= \sum \sigma^2[T_i] + \sum \sum \sigma[T_i, T_j]; i < j \text{-----} 2.26.6 \end{aligned}$$

However:

$$\rho[X_i, X_i] = \sigma[X_i, X_i] / (\sigma[X_i] \sigma[X_i])$$

and;

$$\rho[X_i, X_j] = \sigma[X_i, X_j] / (\sigma[X_i] \sigma[X_j])$$

Therefore:

$$\sigma[X_i, X_i] = \rho[X_i, X_i] \sigma[X_i] \sigma[X_i]$$

and;

$$\sigma[X_i, X_j] = \rho[X_i, X_j] \sigma[X_i] \sigma[X_j]$$

However:

$$\begin{aligned} \sigma[X_i, X_i] &= \sigma[(T_i + E_i)(T_i + E_i)] \\ &= \sigma[T_i, T_i] \\ &= \sigma^2[T_i] \end{aligned}$$

and

$$\begin{aligned} \sigma[X_i, X_j] &= \sigma[(T_i + E_i)(T_j + E_j)] \\ &= \sigma[T_i, T_j] \end{aligned}$$

Consequently:

$$\sigma^2[T_i] = \rho[X_i, X_i] \sigma[X_i] \sigma[X_i]$$

$$\sigma[T_i, T_j] = \rho[X_i, X_j] \sigma[X_i] \sigma[X_j]$$

The component true score variance and the component true score covariance terms in Equation 2.26.6 can therefore be substituted as follows:

$$\begin{aligned} \sigma^2[T] &= \sum \rho[X_i, X_i] \sigma[T_i] \sigma[T_i] + \sum \sum \rho[X_i, X_j] \sigma[T_i] \sigma[T_j] \\ &= \sum \rho[X_i, X_i] \sigma^2[T_i] + \sum \sum \rho[X_i, X_j] \sigma[T_i] \sigma[T_j]; (i < j) \end{aligned} \quad \text{2.26.7}$$

Let the k-item test now be split into two parts containing  $k_1$  and  $k_2$  [ $k_1 \neq k_2$ ;  $k_1 + k_2 = k$ ] items respectively. Items in the first part are denoted by 1, 2, ..., n, m, ...,  $k_1$ ; and the items in the second part are denoted by 1, 2, ..., u, v, ...,  $k_2$ . Equation 2.26.7 can then be rewritten as:

$$\begin{aligned} \sigma^2[T] &= \{\sum \rho[X_n, X_n] \sigma^2[T_n] + \sum \rho[X_u, X_u] \sigma^2[T_u]\} + 2\{\sum \rho[X_n, X_m] \sigma[T_n] \sigma[T_m] \\ &\quad + \sum \rho[X_u, X_v] \sigma[T_u] \sigma[T_v] + \sum \rho[X_n, X_u] \sigma[T_n] \sigma[T_u]\} \\ &= \{k_1 \mu (\rho[X_n, X_n] \sigma^2[T_n]) + k_2 \mu (\rho[X_u, X_u] \sigma^2[T_u])\} \\ &\quad + \{k_1 (k_1 - 1) \mu (\rho[X_n, X_m] \sigma[T_n] \sigma[T_m]) + k_2 (k_2 - 1) \mu (\rho[X_u, X_v] \sigma[T_u] \sigma[T_v]) \\ &\quad + (2k_1 k_2) \mu (\rho[X_n, X_u] \sigma[T_n] \sigma[T_u])\}; (n < m; u < v) \end{aligned} \quad \text{2.26.8}$$

$$\text{Since, } \mu(\rho[X_n, X_n]\sigma^2[T_n]) = \Sigma\rho[X_i, X_i]\sigma^2[T_i]/k_1$$

$$\Sigma\rho[X_i, X_i]\sigma^2[T_i] = k_1\mu(\rho[X_n, X_n]\sigma^2[T_n]) \text{-----} 2.26.9$$

However::

$$\begin{aligned} \mu(\rho[X_n, X_u]\sigma[T_n]\sigma[T_u]) &= \Sigma\rho[X_n, X_u]\sigma[T_n]\sigma[T_u]/(k_1k_2) \\ &= \rho[X_{k1}, X_{k2}]\sigma[T_{k1}]\sigma[T_{k2}]/(k_1k_2) \text{-----} 2.26.10 \end{aligned}$$

$$\text{Since, } \Sigma\rho[X_n, X_u]\sigma[T_n]\sigma[T_u] = \rho[X_{k1}, X_{k2}]\sigma[T_{k1}]\sigma[T_{k2}]$$

Raju [1970] assumes that:

$$\begin{aligned} \mu(\rho[X_n, X_n]\sigma^2[T_n]) &= \mu(\rho[X_u, X_u]\sigma^2[T_u]) \\ &= \mu(\rho[X_n, X_m]\sigma[T_n]\sigma[T_m]) \\ &= \mu(\rho[X_u, X_v]\sigma[T_u]\sigma[T_v]) \\ &= \mu(\rho[X_n, X_u]\sigma[T_n]\sigma[T_u]) \text{-----} 2.26.11 \end{aligned}$$

Substituting Equations 2.26.11 and 2.26.10 in Equation 2.26.8:

$$\begin{aligned} \sigma^2[T] &= \{k_1(\rho[X_1, X_2]\sigma[T_{k1}]\sigma[T_{k2}]/(k_1k_2)) + \{k_2(\rho[X_1, X_2]\sigma[T_{k1}]\sigma[T_{k2}]/(k_1k_2))\} + \\ &\quad \{k_1(k_1-1)(\rho[X_1, X_2]\sigma[T_{k1}]\sigma[T_{k2}]/(k_1k_2))\} + \\ &\quad \{k_2(k_2-1)(\rho[X_1, X_2]\sigma[T_{k1}]\sigma[T_{k2}]/(k_1k_2))\} + \{2k_1k_2(\rho[X_1, X_2]\sigma[T_{k1}]\sigma[T_{k2}]/(k_1k_2))\} \\ &= \rho[X_1, X_2]\sigma[T_{k1}]\sigma[T_{k2}]\{k_1 + k_2 + k_1^2 - k_1 + k_2^2 - k_2 + 2k_1k_2\}/(k_1k_2) \\ &= \rho[X_1, X_2]\sigma[T_{k1}]\sigma[T_{k2}]\{k_1^2 + k_2^2 + 2k_1k_2\}/(k_1k_2) \\ &= \rho[X_1, X_2]\sigma[T_{k1}]\sigma[T_{k2}]\{k_1 + k_2\}^2/(k_1k_2) \text{-----} 2.26.12 \end{aligned}$$

$$\text{Let } \lambda_1 = k_1/k \text{ and } \lambda_2 = k_2/k \text{-----} 2.26.13$$

Therefore:

$$\lambda_1\lambda_2 = k_1k_2/k^2 \text{-----} 2.26.14$$

Therefore:

$$\begin{aligned} 1/\lambda_1\lambda_2 &= k^2/(k_1k_2) \\ &= [k_1 + k_2]^2/(k_1k_2) \text{-----} 2.26.15 \end{aligned}$$

Since by assumption  $k_1 + k_2 = k$



Substituting  $(k_1 + k_2)^2/(k_1 k_2)$  in Equation 2.26.12 with Equation 2.26.15:

$$\sigma^2[T] = \rho[X_1, X_2] \sigma[T_{k1}] \sigma[T_{k2}] (1/\lambda_1 \lambda_2) \text{-----} 2.26.16$$

Therefore:

$$\sigma^2[T]/\sigma^2[X] = \rho[X_1, X_2] \sigma[T_{k1}] \sigma[T_{k2}] / (\sigma^2[X] \lambda_1 \lambda_2) \text{-----} 2.26.17$$

Therefore:

$$\begin{aligned} \rho^2[X, T] &= \rho[X_1, X_2] \sigma[X_1] \sigma[X_2] \\ &= \sigma[X_1, X_2] / (\lambda_1 \lambda_2 \sigma^2[X]) \text{-----} 2.26 \end{aligned}$$

Where,  $\lambda_1 = k_1/(k_1 + k_2)$  and  $\lambda_2 = k_2/(k_1 + k_2)$

Equation 2.26 reduces to the Flanagan coefficient [Equation 2.22] for  $\tau$ -equivalent measurements, if  $k_1 = k_2$  [i.e.  $\lambda_1 = \lambda_2 = 0.50$ ].

Should the length of the two congeneric components, however, not be determinable from some countable characteristic of the components [such as number of items], the Raju coefficient [Equation 2.26] would not apply. Feldt [1975] proposes the coefficient shown as Equation 2.27 for such conditions.

Let  $X$ ,  $X_1$  and  $X_2$  represent the observed scores for the total test and the two components. Let the true scores for the total test and the two components be represented by  $T$ ,  $T_1$  and  $T_2$  and the error scores by  $E$ ,  $E_1$  and  $E_2$ . The component true scores  $T_1$  and  $T_2$  are assumed to constitute unknown proportions [ $\lambda_1$  and  $\lambda_2$ ] of the total test true score  $T$ .

Taking the Equation 2.1 as the basic point of departure:

$$\begin{aligned} X &= X_1 + X_2 \\ &= T + E \text{-----} 2.27.1 \end{aligned}$$

$$\begin{aligned} T &= T_1 + T_2 \\ &= \lambda_1 T + \lambda_2 T \text{-----} 2.27.2 \end{aligned}$$

$$E = E_1 + E_2 \text{-----} 2.27.3$$

In terms of the congeneric measurement model the component observed scores can be broken down into the following parts:

$$X_1 = \lambda_1 T + b_1 + E_1 \text{-----} 2.27.4$$

$$X_2 = \lambda_2 T - b_2 + E_2 \text{-----} 2.27.5$$

Let  $\lambda_1$  and  $\lambda_2$  represent constants [ $\lambda_i > 0$ ;  $\Sigma \lambda_i = 1$ ] which reflect the proportions of the total test true score contributed by the 2 components. . Let  $b_i$  [ $\Sigma b_i = 0$ ] represent constants which provide for differences in component mean scores not attributable to differences in component length.

By utilising Equations 2.27.4 and 2.27.5, the difference in the observed test variance of the two test parts can be written as:

$$\begin{aligned} \sigma^2[X_1] - \sigma^2[X_2] &= \sigma^2[\lambda_1 T + b_1 + E_1] - \sigma^2[\lambda_2 T - b_2 + E_2] \\ &= \{\lambda_1^2 \sigma^2[T] + \sigma^2[b_1] + \sigma^2[E_1] + \lambda_1 \sigma[T, b_1] + \lambda_1 \sigma[T, E_1] + \sigma[b_1, E_1]\} - \\ &\quad \{\lambda_2^2 \sigma^2[T] + \sigma^2[b_2] + \sigma^2[E_2] + \lambda_2 \sigma[T, b_2] + \lambda_2 \sigma[T, E_2] + \sigma[b_2, E_2]\} \\ &= \{\lambda_1^2 \sigma^2[T] + \sigma^2[E_1]\} - \{\lambda_2^2 \sigma^2[T] + \sigma^2[E_2]\} \text{-----} 2.27.6 \end{aligned}$$

$$\text{Since, } \sigma^2[b_i] = \sigma[T, b_i] = \sigma[T, E_i] = \sigma[b_i, E_i] = \sigma[T, E_2] = \sigma[b_2, E_2] = 0$$

Feldt [1975, p. 559] assumes that the “two parts represent differentially shortened versions of the total test” so that:

$$\sigma^2[E_1] = \lambda_1 \sigma^2[E] \text{-----} 2.27.7$$

and:

$$\sigma^2[E_2] = \lambda_2 \sigma^2[E] \text{-----} 2.27.8$$

Substituting Equations 2.27.7 and 2.27.8 in Equation 2.27.6:

$$\begin{aligned} \sigma^2[X_1] - \sigma^2[X_2] &= \lambda_1^2 \sigma^2[T] + \lambda_1 \sigma^2[E] - \lambda_2^2 \sigma^2[T] - \lambda_2 \sigma^2[E] \\ &= \sigma^2[T][\lambda_1^2 - \lambda_2^2] + \sigma^2[E][\lambda_1 - \lambda_2] \\ &= \sigma^2[T][\lambda_1 - \lambda_2][\lambda_1 + \lambda_2] + \lambda[E][\lambda_1 - \lambda_2] \\ &= \sigma^2[T][\lambda_1 - \lambda_2][1] + \sigma^2[E][\lambda_1 - \lambda_2] \\ &= [\lambda_1 - \lambda_2][\sigma^2[T] + \sigma^2[E]] \\ &= [\lambda_1 - \lambda_2] \sigma^2[X] \text{-----} 2.27.9 \end{aligned}$$

Expressing Equation 2.27.9 in terms of the observed test variance of the total test:

$$\begin{aligned}\sigma^2[X_1] - \sigma^2[X_2] / \sigma^2[X] &= ([\lambda_1 - \lambda_2] \sigma^2[X]) / \sigma^2[X] \\ &= [\lambda_1 - \lambda_2] \end{aligned} \quad 2.27.10$$

However by definition:

$$\lambda_1 + \lambda_2 = 1 \quad 2.27.11$$

Isolating  $\lambda_1$  in Equation 2.27.11:

$$\lambda_1 = 1 - \lambda_2 \quad 2.27.12$$

Substituting Equation 2.27.12 in Equation 2.27.10:

$$\begin{aligned}(\sigma^2[X_1] - \sigma^2[X_2]) / \sigma^2[X] &= (1 - \lambda_2) - \lambda_2 \\ &= 1 - 2\lambda_2 \end{aligned} \quad 2.27.13$$

Isolating  $2\lambda_2$  in Equation 2.27.13:

$$\begin{aligned}2\lambda_2 &= 1 - \{(\sigma^2[X_1] - \sigma^2[X_2]) / \sigma^2[X]\} \\ &= \{\sigma^2[X] - (\sigma^2[X_1] - \sigma^2[X_2])\} / \sigma^2[X] \end{aligned} \quad 2.27.14$$

Therefore:

$$\lambda_2 = \{\sigma^2[X] - (\sigma^2[X_1] - \sigma^2[X_2])\} / 2\sigma^2[X] \quad 2.27.15$$

Similarly by substituting  $\lambda_2 = 1 - \lambda_1$  in Equation 2.27.10:

$$\lambda_1 = \{\sigma^2[X] + (\sigma^2[X_1] - \sigma^2[X_2])\} / 2\sigma^2[X] \quad 2.27.17$$

However:

$$\begin{aligned}\sigma[X_1, X_2] &= \sigma[(T_1 + E_1), (T_2 + E_2)] \\ &= \sigma[T_1 T_2 + T_1 E_2 + E_1 T_2 + E_1 E_2] \\ &= \sigma[T_1, T_2] \\ &= \sigma[\lambda_1 T, \lambda_2 T] \\ &= \lambda_1 \lambda_2 \sigma[T, T]\end{aligned}$$

$$= \lambda_1 \lambda_2 \sigma^2[T] \text{-----} 2.27.18$$

Isolating  $\sigma^2[T]$  in Equation 2.27.18 and substituting  $\lambda_1$  and  $\lambda_2$  with Equations 2.27.15 and 2.27.17:

$$\begin{aligned} \sigma^2[T] &= \sigma[X_1, X_2] / \lambda_1 \lambda_2 \\ &= \sigma[X_1, X_2] / (\{\sigma^2[X] - (\sigma^2[X_1] - \sigma^2[X_2])\} / 2\sigma^2[X]) (\{\sigma^2[X] + (\sigma^2[X_1] - \sigma^2[X_2])\} / 2\sigma^2[X]) \\ &= (\sigma[X_1, X_2] 2\sigma^2[X] 2\sigma^2[X]) / \{\sigma^2[X] - (\sigma^2[X_1] - \sigma^2[X_2])\} \{\sigma^2[X] + (\sigma^2[X_1] - \sigma^2[X_2])\} \\ &= (4\sigma[X_1, X_2] \sigma^4[X]) / \{\sigma^4[X] - (\sigma^2[X_1] - \sigma^2[X_2])^2\} \\ &= \sigma^4[X] \{ (4\sigma[X_1, X_2] \sigma^4[X]) / \{\sigma^4[X] - (\sigma^2[X_1] - \sigma^2[X_2])^2\} \} \\ &= (4\sigma[X_1, X_2]) / \{1 - (\sigma^2[X_1] - \sigma^2[X_2])^2 / \sigma^4[X]\} \\ &= (4\sigma[X_1, X_2]) / \{1 - ((\sigma^2[X_1] - \sigma^2[X_2]) / \sigma^2[X])^2\} \text{-----} 2.27.19 \end{aligned}$$

Therefore:

$$\sigma^2[T] / \sigma^2[X] = (4\sigma[X_1, X_2]) / \{\sigma^2[X] - ((\sigma^2[X_1] - \sigma^2[X_2]) / \sigma[X])^2\} \text{-----} 2.27.20$$

Therefore the reliability coefficient can be written as:

$$\begin{aligned} \rho^2[X, T] &= \sigma^2[T] / \sigma^2[X] \\ &= 4\sigma[X_1, X_2] / \{\sigma^2[X] - ((\sigma^2[X_1] - \sigma^2[X_2]) / \sigma[X])^2\} \text{-----} 2.27 \end{aligned}$$

The division of the total instrument into two parallel components is, however, not always possible and, even if it would be possible, not always desirable [Feldt & Brennan, 1989]. Consequently, a need for multi-component reliability coefficients arises. Let  $X_1, X_2, \dots, X_k$  be essentially  $\tau$ -equivalent component measurements with true scores  $T_1, T_2, \dots, T_k$ . Let  $X = \Sigma X_i$  be a composite measurement with true score  $T = \Sigma T_i$ . The Cronbach/Guttman coefficient designed for an instrument divided into two  $\tau$ -equivalent components [Equation 2.21] then generalises to Equation 2.28 as shown below.

Utilising the same initial three steps employed to derive Equation 2.21.

The square of any non-zero difference necessarily will be positive:

$$(\sigma[T_i] - \sigma[T_j])^2 \geq 0 \text{-----} 2.28.1$$

Consequently:

$$\sigma^2[T_i] + \sigma^2[T_j] - 2\sigma[T_i]\sigma[T_j] \geq 0 \text{-----} 2.28.2$$

and thus:

$$\sigma^2[T_i] + \sigma^2[T_j] \geq 2\sigma[T_i]\sigma[T_j] \text{-----} 2.28.3$$

Any positive correlation necessarily must have a value smaller than unity. Therefore:

$$\rho[T_i, T_j] = |\sigma[T_i, T_j]| / (\sigma[T_i]\sigma[T_j]) \leq 1 \text{-----} 2.28.4$$

Cross multiplication of Equation 2.28.4 results in:

$$\begin{aligned} \sigma[T_i]\sigma[T_j] &\geq |\sigma[T_i, T_j]| \\ &\geq \sigma[T_i, T_j] \text{-----} 2.28.5 \end{aligned}$$

Multiplying both sides of Equation 2.28.5 by 2 results in:

$$2\sigma[T_i]\sigma[T_j] \geq 2\sigma[T_i, T_j] \text{-----} 2.28.6$$

Combining Equations 2.28.6 and 2.28.3 results in:

$$\sigma^2[T_i] + \sigma^2[T_j] \geq 2\sigma[T_i, T_j] \text{-----} 2.28.7$$

Thus, when summing for  $i \neq j$ ,

$$\Sigma\Sigma(\sigma^2[T_i] + \sigma^2[T_j]) \geq 2\Sigma\Sigma\sigma[T_i, T_j] \text{-----} 2.28.8$$

However:

$$\begin{aligned} \Sigma\Sigma(\sigma^2[T_i] + \sigma^2[T_j]) &= \Sigma\Sigma\sigma^2[T_i] + \Sigma\Sigma\sigma^2[T_j] \\ &= k\Sigma\sigma^2[T_i] + k\Sigma\sigma^2[T_j] \\ &= 2k\Sigma\sigma^2[T_i] \text{-----} 2.28.9 \end{aligned}$$

and also:

$$\begin{aligned} \Sigma\Sigma(\sigma^2[T_i] + \sigma^2[T_j]) &= (i=j)\Sigma\Sigma(\sigma^2[T_i] + \sigma^2[T_j]) + (i \neq j)\Sigma\Sigma(\sigma^2[T_i] + \sigma^2[T_j]) \\ &= 2\Sigma\sigma^2[T_i] + (i \neq j)\Sigma\Sigma(\sigma^2[T_i] + \sigma^2[T_j]) \text{-----} 2.28.10 \end{aligned}$$

Comparing Equations 2.28.9 and 2.28.10 results in:



$$2k\sigma^2[T_i] = 2\sum\sigma^2[T_i] + (i \neq j)\sum(\sigma^2[T_i] + \sigma^2[T_j]) \text{-----} 2.28.11$$

Isolating the term  $(i \neq j)\sum(\sigma^2[T_i] + \sigma^2[T_j])$  in Equation 2.28.11 and subsequently factoring out  $2\sum\sigma^2[T_i]$ , results in:

$$\begin{aligned} (i \neq j)\sum(\sigma^2[T_i] + \sigma^2[T_j]) &= 2k\sigma^2[T_i] - 2\sum\sigma^2[T_i] \\ &= 2\sum\sigma^2[T_i](k - 1) \text{-----} 2.28.12 \end{aligned}$$

Substituting for  $(i \neq j)\sum(\sigma^2[T_i] + \sigma^2[T_j])$  in Equation 2.28.8 with Equation 2.28.12 results in:

$$2\sum\sigma^2[T_i](k - 1) \geq 2\sum\sigma[T_i, T_j] \text{-----} 2.28.13$$

Multiplying both sides of Equation 2.28.14 with  $(1/(k-1))$  results in:

$$\sum\sigma^2[T_i] \geq \sum\sigma[T_i, T_j]/(k - 1) \text{-----} 2.28.14$$

However, since  $X$  is composite measure with true score  $T = \sum T_i$ , the true score variance can be analysed as follows:

$$\begin{aligned} \sigma^2[T] &= \sigma^2[\sum T_i] \\ &= \sum\sigma^2[T_i] + \sum\sigma[T_i, T_j] \text{-----} 2.28.15 \end{aligned}$$

Isolating the term  $\sum\sigma^2[T_i]$  in Equation 2.28.15 results in:

$$\sum\sigma^2[T_i] = \sigma^2[T] - \sum\sigma[T_i, T_j] \text{-----} 2.28.16$$

Substituting  $\sum\sigma^2[T_i]$  in Equation 2.28.24 with Equation 2.28.16 results in:

$$\sigma^2[T] - \sum\sigma[T_i, T_j] \geq \sum\sigma[T_i, T_j]/(k - 1) \text{-----} 2.28.17$$

Isolating the term  $\sigma^2[T]$  in Equation 2.28.17 and multiplying the right-hand term in Equation 2.28.17 with  $(k-1)/(k-1)$ , results in:

$$\begin{aligned} \sigma^2[T] &\geq \sum\sigma[T_i, T_j]/(k - 1) + \sum\sigma[T_i, T_j] \\ &\geq \sum\sigma[T_i, T_j]/(k - 1) + \sum\sigma[T_i, T_j](k - 1)/(k - 1) \\ &\geq \sum\sigma[T_i, T_j]\{(1+k-1)/(k-1)\} \\ &\geq \{k/(k-1)\}\sum\sigma[T_i, T_j] \text{-----} 2.28.18 \end{aligned}$$

However, since  $X$  is composite measure with observed score  $X = \sum X_i$ , the observed test score variance can be analysed as follows:

$$\begin{aligned}\sigma^2[X] &= \sigma^2[\sum X_i] \\ &= \sum \sigma^2[X_i] + \sum \sum \sigma[X_i, X_j] \text{-----} 2.28.19\end{aligned}$$

In accordance with Equation 2.1 the component measure  $X_i = T_i + E_i$ . The covariance term in Equation 2.28.19 can therefore be analysed as follows:

$$\begin{aligned}\sum \sum \sigma[X_i, X_j] &= \sum \sum \sigma[(T_i + E_i), (T_j + E_j)] \\ &= \sum \sum \sigma[T_i, T_j] + \sum \sigma[T_i, E_j] + \sum \sum \sigma[E_i, T_j] + \sum \sum \sigma[E_i, E_j] \\ &= \sum \sum \sigma[T_i, T_j] \text{-----} 2.28.20\end{aligned}$$

$$\text{Since } \sum \sum \sigma[T_i, E_j] = \sum \sum \sigma[E_i, T_j] = \sum \sum \sigma[E_i, E_j] = 0$$

Substituting Equation 2.28.20 into Equation 2.28.19 results in:

$$\sigma^2[X] = \sum \sigma^2[X_i] + \sum \sum \sigma[T_i, T_j] \text{-----} 2.28.21$$

Isolating the covariance term in Equation 2.29.21 results in:

$$\sum \sum \sigma[T_i, T_j] = \sigma^2[X] - \sum \sigma^2[X_i] \text{-----} 2.28.22$$

Substituting Equation 2.28.22 into Equation 2.28.18 results in:

$$\begin{aligned}\sigma^2[T] &\geq \{k/(k-1)\} \sum \sum \sigma[T_i, T_j] \\ &\geq \{k/(k-1)\} (\sigma^2[X] - \sum \sigma^2[X_i]) \text{-----} 2.28.23\end{aligned}$$

Therefore, when multiplying Equation 2.28.23 with  $(1/\sigma^2[X])$ :

$$\begin{aligned}\sigma^2[T]/\sigma^2[X] &\geq (1/\sigma^2[X]) (k/(k-1)) (\sigma^2[X] - \sum \sigma^2[X_i]) \\ &\geq (k/(k-1)) (1 - \sum \sigma^2[X_i]/\sigma^2[X]) \text{-----} 2.28.24\end{aligned}$$

Write:

$$\alpha = (k/(k-1)) (1 - \sum \sigma^2[X_i]/\sigma^2[X]) \text{-----} 2.28$$

Equation 2.28 is generally known as coefficient  $\alpha$  [Cronbach, 1951] as it would apply to a  $k$  component composite or  $\lambda_3$  [Guttman, 1954]. Coefficient  $\alpha$  will return a lower bound of the coefficient of precision unless the  $k$  components may be considered essentially tau equivalent, in which case Equation 2.28 holds as an equality [Lord & Novick, 1968; Novick & Lewis, 1967; Zimmerman, Zumbo & Lalonde, 1993].

If the  $k$  component measurements,  $X_1, X_2, \dots, X_k$ , are parallel measurements, Equation 2.28 reduces to the generalised Spearman-Brown formula [Equation 2.20], as shown by Equation 2.29.

If the  $k$  component measurements,  $X_1, X_2, \dots, X_k$ , are parallel measurements, the  $k$  component variances will be equal. Equation 2.28 can thus be simplified as follows:

$$\begin{aligned}\alpha &= (k/(k-1))(1 - \Sigma\sigma^2[X_i]/\sigma^2[X]) \\ &= (k/(k-1))(1 - k\sigma^2[X_i]/\sigma^2[X])\end{aligned}\quad 2.29.1$$

Because the test score is a linear composite of parallel measurements and because all components constitute parallel measurements and thus all  $\rho[X_i X_j]$  are equal and  $\sigma[X_i] = \sigma[X_j]$  for all  $i$  and  $j$ , the test score variance can be analysed as follows:

$$\begin{aligned}\sigma^2[X] &= \sigma^2[\Sigma X_i] \\ &= \Sigma\sigma^2[X_i] + \Sigma\Sigma\sigma[X_i X_j] \\ &= \Sigma\sigma^2[X_i] + \Sigma\Sigma\rho[X_i X_j]\sigma[X_i]\sigma[X_j] \\ &= k\sigma^2[X_i] + k(k-1)\rho[X_i X_j]\sigma^2[X_i]\end{aligned}\quad 2.29.2$$

Equation 2.29.2 can be simplified further by factoring out the term  $k\sigma^2[X_i]$ :

$$\sigma^2[X] = k\sigma^2[X_i]\{1 + (k-1)\rho[X_i X_j]\}\quad 2.29.3$$

Substituting Equation 2.29.3 into Equation 2.29.1 results in:

$$\begin{aligned}\alpha &= (k/(k-1))[1 - k\sigma^2[X_i]/(k\sigma^2[X_i]\{1 + (k-1)\rho[X_i X_j]\})] \\ &= (k/(k-1))(\{k\sigma^2[X_i]\{1 + (k-1)\rho[X_i X_j]\} - k\sigma^2[X_i]\}/(k\sigma^2[X_i]\{1 + (k-1)\rho[X_i X_j]\})) \\ &= (k/(k-1))\{1 + (k-1)\rho[X_i X_j] - 1\}/(1 + (k-1)\rho[X_i X_j]) \\ &= (k/(k-1))\{(k-1)\rho[X_i X_j]\}/(1 + (k-1)\rho[X_i X_j]) \\ &= k\rho[X_i X_j]/(1 + (k-1)\rho[X_i X_j])\end{aligned}\quad 2.29.4$$

Write:

$$\rho_{ttX} = k\rho[X_i, X_j] / (1 + (k-1)\rho[X_i, X_j]) \text{-----} 2.29$$

If the individual item components comprising the  $k$  component composite measurement are binary/dichotomous random variables with possible values of zero and one and associated probabilities of  $q_i = [1-p_i]$  and  $p_i$ , coefficient  $\alpha$  reduces to the Kuder-Richardson formula 20 [KR20] shown as Equation 2.30 [Cronbach, 1951; Kuder, 1991; Kuder & Richardson, 1937; Lord & Novick, 1968].

Using Equation 2.28 as basic point of departure

$$\alpha = (k/(k-1))(1 - \Sigma\sigma^2[X_i]/\sigma^2[X]) \text{-----} 2.30.1$$

The observed score variance can be expressed as:

$$\sigma^2[X_i] = E[X_i^2] - E[X_i]^2 \text{-----} 2.30.2$$

But for binary items the expected value of  $X^2$  equals the expected value of  $X$ :

$$E[X_i^2] = E[X_i] \text{-----} 2.30.3$$

Assuming binary items and applying Equation 2.30.3 to Equation 2.30.2 results in:

$$\begin{aligned} \sigma^2[X_i] &= E[X_i] - E[X_i]^2 \\ &= E[X_i](1 - E[X_i]) \text{-----} 2.30.4 \end{aligned}$$

Write:

$$E[X_i] = p_i$$

Write:

$$q_i = [1-p_i]$$

Equation 2.30.4 can then be written as:

$$\begin{aligned} \sigma^2[X_i] &= p_i - p_i^2 \\ &= p_i[1 - p_i] \\ &= p_i q_i \text{-----} 2.30.5 \end{aligned}$$

Substituting  $\sigma^2[X_i]$  in Equation 2.30.1 with Equation 2.30.5 results in:

$$\alpha = KR20 = (k/(k-1))(1 - \sum p_i q_i / \sigma^2[X]) \text{-----} 2.30$$

If the assumption can be made that the  $k$  items comprising the test have identical difficulty values  $p_i$ , the Kuder-Richardson formula 20 reduces to the Kuder-Richardson formula 21 [KR21] shown as Equation 2.31 [Kuder, 1991; Kuder & Richardson, 1937; Lord & Novick, 1968].

The variance in the  $k$  item difficulty values  $p_i$  can be obtained through:

$$\sigma^2[p] = E[p^2] - E[p]^2 \text{-----} 2.31.1$$

Isolating  $E[p]^2$  in Equation 2.31.1:

$$E[p^2] = \sigma^2[p] + E[p]^2 \text{-----} 2.31.2$$

However:

$$\begin{aligned} \mu[pq] &= E[pq] \\ &= E[p(1-p)] \\ &= E[p] - E[p^2] \text{-----} 2.31.3 \end{aligned}$$

Substituting Equation 2.31.2 into Equation 2.31.3:

$$\begin{aligned} \mu[pq] &= E[p] - (\sigma^2[p] + E[p]^2) \\ &= E[p] - \sigma^2[p] - E[p]^2 \\ &= E[p](1 - E[p]) - \sigma^2[p] \\ &= \mu[p]\mu[q] - \sigma^2[p] \text{-----} 2.31.4 \end{aligned}$$

If identical difficulty values are assumed for all  $k$  items then  $\sigma^2[p] = 0$ , and Equation 2.31.4 reduces to:

$$\mu[pq] = \mu[p]\mu[q] \text{-----} 2.31.5$$

Simultaneously:

$$\mu[pq] = \sum p_i q_i / k \text{-----} 2.31.6$$

Cross multiplying Equation 2.31.6 and substituting  $\mu[pq]$  with Equation 2.31.5:



$$\begin{aligned}\Sigma p_i q_i &= k\mu[pq] \\ &= k\mu[p]\mu[q] \text{-----} 2.31.7\end{aligned}$$

Substituting in Equation 2.30 with Equation 2.31.7:

$$\alpha = KR20 = (k/(k-1))(1 - k\mu[p]\mu[q]/\sigma^2[X]) \text{-----} 2.31.8$$

Consequently, if identical difficulty values are assumed for all k items:

$$KR20 = KR21 = (k/(k-1))(1 - k\mu[p]\mu[q]/\sigma^2[X]) \text{-----} 2.31$$

To the extent that the assumption of identical  $p_i$ -values is in fact not satisfied, the KR21 is negatively biased. Tucker [1949] presented an equation to correct for the underestimation of KR20 shown as Equation 2.32.

Using Equation 2.26 as point of departure:

$$KR20 = (k/(k-1))(1 - \Sigma p_i q_i / \sigma^2[X])$$

According to Equation 2.31.4 and Equation 2.31.6:

$$\begin{aligned}\mu[p_i q_i] &= \mu[p_i]\mu[q_i] - \sigma^2[p_i] \\ &= \Sigma p_i q_i / k \text{-----} 2.32.1\end{aligned}$$

Thus, if non-identical difficulty values are assumed for all k items:

$$\begin{aligned}\Sigma p_i q_i &= k\mu[p_i q_i] \\ &= k[\mu[p_i]\mu[q_i] - \sigma^2[p_i]] \\ &= k\mu[p_i]\mu[q_i] - k\sigma^2[p_i] \text{-----} 2.32.2\end{aligned}$$

Thus, by inserting Equation 2.32.2 into Equation 2.26:

$$\begin{aligned}KR20 &= (k/(k-1))(1 - \{k\mu[p_i]\mu[q_i] - k\sigma^2[p_i]\} / \sigma^2[X]) \\ &= (k/(k-1))(1 - \{k\mu[p_i]\mu[q_i] / \sigma^2[X]\}) - ((k/(k-1))(k\sigma^2[p_i]) / \sigma^2[X]) \\ &= (k/(k-1))(1 - k\mu[p_i]\mu[q_i] / \sigma^2[X]) - ((k^2\sigma^2[p_i]) / (k-1)\sigma^2[X]) \text{-----} 2.32.3\end{aligned}$$

Consequently, if non-identical difficulty values are assumed for all k items the following correction applies:

$$KR_{20} = KR_{21} + k^2\sigma^2[p_i]/(k-1)\sigma^2[X] \text{-----} 2.32$$

Another reliability coefficient based on more than two  $\tau$ -equivalent components was suggested by Guttman [1945]. Like Cronbach's  $\alpha$  and the Kuder-Richardson formula, Guttman's  $\lambda_2$  returns a lower bound to the reliability of a measure [Cronbach, 1951; Guttman, 1954; Kuder & Richardson, 1937; ten Berge & Zegers, 1978]. The Guttman reliability coefficient  $\lambda_2$  is shown as Equation 2.33.

$$\begin{aligned} \lambda_2 &= (\Sigma\Sigma\sigma[X_i, X_j] + \{k/(k-1)\Sigma\Sigma(i \neq j)\sigma^2[X_i, X_j]\}^{1/2})/\sigma^2[X] \\ &= ((\sigma^2[X] - \Sigma\sigma^2[X_i]) + \{k/(k-1)\Sigma\Sigma(i \neq j)\sigma^2[X_i, X_j]\}^{1/2})/\sigma^2[X] \text{-----} 2.33 \end{aligned}$$

If Equation 2.33 does provide a lower bound to the reliability coefficient, then:

$$\begin{aligned} \lambda_2 &\leq \rho^2[X, T] \\ &\leq \sigma^2[T]/\sigma^2[X] \text{-----} 2.33.1 \end{aligned}$$

Substituting Equation 2.33 in Equation 2.33.1 and isolating  $\sigma^2[T]$ :

$$\begin{aligned} \sigma^2[T] &\geq \{((\sigma^2[X] - \Sigma\sigma^2[X_i]) + \{k/(k-1)\Sigma\Sigma(i \neq j)\sigma^2[X_i, X_j]\}^{1/2})/\sigma^2[X]\}\sigma^2[X] \\ &\geq \Sigma\Sigma\sigma[X_i, X_j] + \{k/(k-1)\Sigma\Sigma\sigma^2[X_i, X_j]\}^{1/2} \text{-----} 2.33.2 \end{aligned}$$

Rearranging the terms in Equation 2.33.2:

$$\sigma^2[T] - \Sigma\Sigma\sigma[X_i, X_j] \geq (k/(k-1)\Sigma\Sigma\sigma^2[X_i, X_j])^{1/2} \text{-----} 2.33.3$$

However:

$$\begin{aligned} \sigma^2[T] &= \sigma^2[\Sigma T_i] \\ &= \Sigma\sigma^2[T_i] + \Sigma\Sigma\sigma[X_i, X_j] \text{-----} 2.33.4 \end{aligned}$$

Isolating the term  $\Sigma\sigma^2[T_i]$  in Equation 2.33.4:

$$\Sigma\sigma^2[T_i] = \sigma^2[T] - \Sigma\Sigma\sigma[X_i, X_j] \text{-----} 2.33.5$$

Therefore, by substituting Equation 2.33.5 in Equation 2.33.3 and if Equation 2.33 does provide a lower bound to the reliability coefficient, then:

$$\Sigma\sigma^2[T_i] \geq (k/(k-1)\Sigma\Sigma\sigma^2[X_i, X_j])^{1/2} \text{-----} 2.33.6$$

Furthermore, because any non-zero squared difference necessarily must be positive:

$$(\sigma^2[T_i] - \sigma^2[T_j])^2 \geq 0 \text{-----} 2.33.7$$

Therefore:

$$\sigma^4[T_i] + \sigma^4[T_j] - 2\sigma^2[T_i]\sigma^2[T_j] \geq 0 \text{-----} 2.33.8$$

Therefore:

$$\sigma^4[T_i] + \sigma^4[T_j] \geq 2\sigma^2[T_i]\sigma^2[T_j] \text{-----} 2.33.9$$

However, since  $-1 \leq \rho[T_i, T_j] \leq 1$ :

$$\rho^2[T_i, T_j] = \sigma^2[T_i, T_j] / (\sigma^2[T_i]\sigma^2[T_j]) \leq 1 \text{-----} 2.33.10$$

Equation 2.33.10 can only be true if:

$$\sigma^2[T_i]\sigma^2[T_j] \geq \sigma^2[T_i, T_j] \text{-----} 2.33.11$$

Multiplying Equation 2.33.11 by 2:

$$2\sigma^2[T_i]\sigma^2[T_j] \geq 2\sigma^2[T_i, T_j] \text{-----} 2.33.12$$

Combining Equations 2.33.9 and 2.33.12:

$$\sigma^4[T_i] + \sigma^4[T_j] \geq 2\sigma^2[T_i, T_j] \text{-----} 2.33.13$$

Thus, if summing for  $i \neq j$ :

$$\Sigma \Sigma [\sigma^4[T_i] + \sigma^4[T_j]] \geq 2 \Sigma \Sigma \sigma^2[T_i, T_j] \text{-----} 2.33.14$$

However:

$$\begin{aligned} \Sigma \Sigma [\sigma^4[T_i] + \sigma^4[T_j]] &= \Sigma \Sigma \sigma^4[T_i] + \Sigma \Sigma \sigma^4[T_j] \\ &= k\sigma^4[T_i] + k\Sigma \sigma^4[T_j] \\ &= 2k\Sigma \sigma^4[T_i] \text{-----} 2.33.15 \end{aligned}$$

The left hand term in Equation 2.33.15 can, however, also be written as:

$$\begin{aligned}\Sigma\Sigma[\sigma^4[T_i] + \sigma^4[T_j]] &= \Sigma\Sigma(i=j)(\sigma^4[T_i] + \sigma^4[T_j]) + \Sigma\Sigma(i\neq j)(\sigma^4[T_i] + \sigma^4[T_j]) \\ &= 2\Sigma\sigma^4[T_i] + \Sigma\Sigma(i\neq j)(\sigma^4[T_i] + \sigma^4[T_j])\end{aligned}\quad 2.33.16$$

Substituting Equation 2.33.15 in Equation 2.33.16:

$$2k\Sigma\sigma^4[T_i] = 2\Sigma\sigma^4[T_i] + \Sigma\Sigma(i\neq j)(\sigma^4[T_i] + \sigma^4[T_j])\quad 2.33.17$$

Isolating  $\Sigma\Sigma(i\neq j)(\sigma^4[T_i] + \sigma^4[T_j])$  in Equation 2.33.17 and factoring out the term  $2\Sigma\sigma^4[T_i]$ :

$$\begin{aligned}\Sigma\Sigma(i\neq j)(\sigma^4[T_i] + \sigma^4[T_j]) &= 2k\Sigma\sigma^4[T_i] - 2\Sigma\sigma^4[T_i] \\ &= 2\Sigma\sigma^4[T_i](k - 1)\end{aligned}\quad 2.33.18$$

Combining Equations 2.33.13 and 2.44.18:

$$2\Sigma\sigma^4[T_i](k - 1) \geq 2\Sigma\Sigma\sigma^2[T_i, T_j]\quad 2.33.19$$

Isolating the term  $\Sigma\sigma^4[T_i]$  in Equation 2.33.19:

$$\Sigma\sigma^4[T_i] \geq \Sigma\Sigma\sigma^2[T_i, T_j]/(k - 1)\quad 2.33.20$$

However:

$$(\Sigma\sigma^2[T_i])^2 = \Sigma\sigma^4[T_i] + \Sigma\Sigma\sigma^2[T_i]\sigma^2[T_j]\quad 2.33.21$$

Therefore, by rearranging the terms of Equation 2.33.21 and by utilising Equation 2.33.11

$$\begin{aligned}(\Sigma\sigma^2[T_i])^2 - \Sigma\sigma^4[T_i] &= \Sigma\Sigma\sigma^2[T_i]\sigma^2[T_j] \\ &\geq \Sigma\Sigma\sigma^2[T_i, T_j] \\ &\geq \Sigma\Sigma\sigma^2[X_i, X_j]\end{aligned}\quad 2.33.22$$

Since  $\Sigma\Sigma\sigma^2[X_i, X_j] = \Sigma\Sigma\sigma^2[T_i, T_j]$

Isolating the term  $(\Sigma\sigma^2[T_i])^2$  in Equation 2.33.22:

$$(\Sigma\sigma^2[T_i])^2 \geq \Sigma\sigma^4[T_i] + \Sigma\Sigma\sigma^2[X_i, X_j]\quad 2.33.23$$

Substituting Equation 2.33.20 in Equation 2.33.23 and multiplying the term  $\Sigma\Sigma\sigma^2[X_i, X_j]$  with  $(k-1)/(k-1)$ :

$$\begin{aligned}
 (\Sigma\sigma^2[T_i])^2 &\geq \Sigma\Sigma\sigma^2[T_i, T_j]/(k-1) + \Sigma\Sigma\sigma^2[X_i, X_j] \\
 &\geq \Sigma\Sigma\sigma^2[T_i, T_j]/(k-1) + (k-1)\Sigma\Sigma\sigma^2[X_i, X_j]/(k-1) \\
 &\geq \Sigma\Sigma\sigma^2[X_i, X_j](1+k-1)/(k-1) \\
 &\geq \Sigma\Sigma\sigma^2[X_i, X_j](k/(k-1)) \text{-----} 2.33.24
 \end{aligned}$$

Since  $\Sigma\Sigma\sigma^2[X_i, X_j] = \Sigma\Sigma\sigma^2[T_i, T_j]$

Consequently, since Equation 2.33.24 is identical to Equation 2.33.6 derived from Guttman's  $\lambda_2$  formula:

$$\begin{aligned}
 \lambda_2 &= (\Sigma\Sigma\sigma[X_i, X_j] + \{k/(k-1)\Sigma\Sigma\sigma^2[X_i, X_j]\}^{1/2})/\sigma^2[X] \\
 &\leq \sigma^2[T]/\sigma^2[X] \text{-----} 2.33
 \end{aligned}$$

According to Feldt and Brennan [1989], Equation 2.33 equals the Cronbach  $\alpha$  in the parameter, if the assumption of essential  $\tau$ -equivalence holds for all components. The probability of exact equality on sample data, however, seems low [Feldt & Brennan, 1989]. Lord and Novick [1968] recommend the use of Guttman's  $\lambda_2$  when some of the test items have negative inter-correlations.

Gilmer and Feldt [1983] derived a reliability coefficient for multiple congeneric measurements. The Gilmer and Feldt reliability coefficient is shown as Equation 2.34.

Assume  $w$  congeneric components. Let the lengths of the components be  $k_1, \dots, k_i, \dots, k_w$ ;  $k$  representing the number of items in the component. Therefore, in accordance with the classical measurement model:

$$X = [T + E]$$

$$X = [X_1 + X_2 + \dots + X_w]$$

$$T = [T_1 + T_2 + \dots + T_w]$$

$$E = [E_1 + E_2 + \dots + E_w]$$

Let  $\lambda_j$  represent constants [ $\lambda_j > 0$ ;  $\Sigma\lambda_j = 1$ ] which reflect the proportions of the total test true score contributed by the  $w$  components. Let  $b_j$  [ $\Sigma b_j = 0$ ] represent constants which provide for differences in

component mean scores not attributable to differences in component length. In terms of the congeneric measurement model the component observed scores can be broken down into the following parts:

$$\begin{aligned} X_j &= T_j + E_j \\ &= \lambda_j T + b_j + E_j; j = 1, 2, \dots, w. \end{aligned} \quad \text{-----} \quad 2.34.1$$

The observed total test score variance can be analysed as follows:

$$\begin{aligned} \sigma^2[X] &= \sigma^2[X_1 + X_2 + \dots + X_k] \\ &= \Sigma \sigma^2[X_i] + \Sigma \Sigma \sigma[X_i, X_j] [i \neq j] \end{aligned} \quad \text{-----} \quad 2.34.2$$

Isolating the covariance term:

$$\Sigma \Sigma \sigma[X_i, X_j] = \sigma^2[X] - \Sigma \sigma^2[X_i] \quad \text{-----} \quad 2.34.3$$

From Equation 2.34.1, however, the covariance term in Equation 2.34.3 can be rewritten as

$$\begin{aligned} \sigma[X_i, X_j] &= \sigma[(T_i + E_i), (T_j + E_j)] \\ &= \sigma[T_i, T_j] + \sigma[T_i, E_j] + \sigma[E_i, T_j] + \sigma[E_i, E_j] \\ &= \sigma[\lambda_i T, \lambda_j T] \\ &= \lambda_i \lambda_j \sigma^2[T] \end{aligned} \quad \text{-----} \quad 2.34.4$$

$$\text{Since } \sigma[T, E_j] = \sigma[E_i, T] = \sigma[E_i, E_j] = 0$$

Substituting Equation 2.34.4 into Equation 2.34.3:

$$\Sigma \Sigma \lambda_i \lambda_j \sigma^2[T] = \sigma^2[X] - \Sigma \sigma^2[X_i] \quad \text{-----} \quad 2.34.5$$

Isolating the total test true score variance in Equation 2.34.5

$$\sigma^2[T] = (1 / \Sigma \Sigma \lambda_i \lambda_j) (\sigma^2[X] - \Sigma \sigma^2[X_i]) \quad \text{-----} \quad 2.34.6$$

However:

$$\begin{aligned} \Sigma \lambda_i^2 &= [\lambda_1 + \lambda_2 + \dots + \lambda_k]^2 \\ &= 1^2 \\ &= \Sigma \lambda_i^2 + \Sigma \Sigma \lambda_i \lambda_j \end{aligned}$$



$$= 1 \text{-----} 2.34.7$$

Isolating the term  $\sum \lambda_i \lambda_j$  in Equation 2.34.7:

$$\sum \lambda_i \lambda_j = 1 - \sum \lambda_i^2 \text{-----} 2.34.8$$

Substituting Equation 2.34.8 into Equation 2.34.6

$$\sigma^2[T] = (1/(1 - \sum \lambda_i^2))(\sigma^2[X] - \sum \sigma^2[X_i])$$

Therefore:

$$\begin{aligned} \rho_{ttX} &= \sigma^2[T]/\sigma^2[X] = \{1/(1 - \sum \lambda_i^2)\} \{(\sigma^2[X] - \sum \sigma^2[X_i])/\sigma^2[X]\} \\ &= \{1/(1 - \sum \lambda_i^2)\} \{1 - \sum \sigma^2[X_i]/\sigma^2[X]\} \text{-----} 2.34 \end{aligned}$$

The Gilmer and Feldt coefficient will equal the Cronbach  $\alpha$  if the  $w$  components are of equal length so that all  $\lambda_i$  equal  $1/w$  as shown in Equation 2.35.

Using Equation 2.34 as the point of departure and assuming the  $w$  components to be of equal length so that all  $\lambda_i$  equal  $1/w$ :

$$\begin{aligned} \rho_{ttX} &= (\sigma^2[X] - \sum \sigma^2[X_i]) / (1 - \sum \lambda_i^2) \sigma^2[X] \\ &= (\sigma^2[X] - \sum \sigma^2[X_i]) / (1 - w/w^2) \sigma^2[X] \\ &= (\sigma^2[X] - \sum \sigma^2[X_i]) / \{(w^2 - w)/w^2\} \sigma^2[X] \\ &= (\sigma^2[X] - \sum \sigma^2[X_i]) / \{(w(w - 1))/w^2\} \sigma^2[X] \\ &= (\sigma^2[X] - \sum \sigma^2[X_i]) / \{(w - 1)/w\} \sigma^2[X] \\ &= \{w/(w-1)\} (1 - \sum \sigma^2[X_i] / \sigma^2[X]) \text{-----} 2.35 \end{aligned}$$

Kristof [1974] derived a method of estimating the reliability coefficient for a test divided into three congeneric components of unknown length shown in Equation 2.36. Kristof [1974] assumes that a test has been split into three content-homogeneous components with observed scores  $X$ ,  $X_1$ ,  $X_2$  and  $X_3$ , true scores  $T$ ,  $T_1$ ,  $T_2$  and  $T_3$ , and error scores  $E$ ,  $E_1$ ,  $E_2$  and  $E_3$ . In accordance with the classical test model, Kristof [1974] furthermore assumes that these random variables satisfy the equations shown below.

$$X = T + E$$

$$X = X_1 + X_2 + X_3$$

$$T = T_1 + T_2 + T_3$$

$$E = E_1 + E_2 + E_3$$

Let  $\lambda_j$  represent constants [ $\lambda_j > 0$ ;  $\sum \lambda_j = 1$ ] which reflect the proportions of the total test true score contributed by the  $w$  components. Let  $b_j$  [ $\sum b_j = 0$ ] represent constants which provide for differences in component mean scores not attributable to differences in component length. In terms of the congeneric measurement model the three component observed scores can be broken down into the following parts:

$$\begin{aligned} X_j &= T_j + E_j; j = 1, 2, 3 \\ &= \lambda_j T + b_j + E_j \end{aligned} \quad \text{-----} \quad 2.36.1$$

Utilising Equation 2.36.1, the component observe score covariance can be analysed as follows:

$$\begin{aligned} \sigma[X_1, X_2] &= \sigma[(T_1 + E_1)(T_2 + E_2)] \\ &= \sigma[T_1 T_2 + T_1 E_2 + E_1 T_2 + E_1 E_2] \\ &= \sigma[T_1, T_2] \\ &= \sigma[\lambda_1 T, \lambda_2 T] \\ &= \lambda_1 \lambda_2 \sigma^2[T] \end{aligned} \quad \text{-----} \quad 2.36.2$$

Similarly it can be shown that:

$$\sigma[X_1, X_3] = \lambda_1 \lambda_3 \sigma^2[T] \quad \text{-----} \quad 2.36.3$$

Similarly it can be shown that:

$$\sigma[X_2, X_3] = \lambda_2 \lambda_3 \sigma^2[T] \quad \text{-----} \quad 2.36.4$$

Combining Equations 2.36.2 and 2.36.3 therefore implies:

$$\begin{aligned} \sigma[X_1, X_2] / \sigma[X_1, X_3] &= \lambda_1 \lambda_2 \sigma^2[T] / (\lambda_1 \lambda_3 \sigma^2[T]) \\ &= \lambda_2 / \lambda_3 \end{aligned} \quad \text{-----} \quad 2.36.5$$

Similarly, by combining Equations 2.36.2 and 2.36.4 it can be shown that:

$$\sigma[X_1, X_2] / \sigma[X_2, X_3] = \lambda_1 / \lambda_3 \quad \text{-----} \quad 2.36.6$$

By combining Equations 2.36.3 and 2.36.4 it can be shown that:

$$\sigma[X_1, X_3]/\sigma[X_2, X_3] = \lambda_1/\lambda_2 \text{-----} 2.36.7$$

By combining Equations 2.36.3 and 2.36.2 it can be shown that:

$$\sigma[X_1, X_3]/\sigma[X_1, X_2] = \lambda_3/\lambda_2 \text{-----} 2.36.8$$

By combining Equations 2.36.4 and 2.36.3 it can be shown that:

$$\sigma[X_2, X_3]/\sigma[X_1, X_3] = \lambda_2/\lambda_1 \text{-----} 2.36.9$$

By combining Equations 2.36.4 and 2.36.2 it can be shown that:

$$\sigma[X_2, X_3]/\sigma[X_1, X_2] = \lambda_3/\lambda_1 \text{-----} 2.36.10$$

Because  $(\lambda_1 + \lambda_2 + \lambda_3) = 1$ , it is possible to write:

$$\begin{aligned} [1/\lambda_1] &= (\lambda_1 + \lambda_2 + \lambda_3)/\lambda_1 \\ &= (1 + \lambda_2/\lambda_1 + \lambda_3/\lambda_1) \text{-----} 2.36.11 \end{aligned}$$

Taking the inverse of Equation 2.36.11 and substituting Equations 2.36.9 and 2.36.10:

$$\begin{aligned} (1/\lambda_1)^{-1} &= (1 + \lambda_2/\lambda_1 + \lambda_3/\lambda_1)^{-1} \\ &= (1 + \sigma[X_2, X_3]/\sigma[X_1, X_3] + \sigma[X_2, X_3]/\sigma[X_1, X_2])^{-1} \\ &= \lambda_1 \text{-----} 2.36.12 \end{aligned}$$

Similarly, it could be shown that:

$$\begin{aligned} \lambda_2 &= (1 + \lambda_1/\lambda_2 + \lambda_3/\lambda_2)^{-1} \\ &= (1 + \sigma[X_1, X_3]/\sigma[X_2, X_3] + \sigma[X_1, X_3]/\sigma[X_1, X_2])^{-1} \text{-----} 2.36.13 \end{aligned}$$

and that:

$$\begin{aligned} \lambda_3 &= (1 + \lambda_1/\lambda_3 + \lambda_2/\lambda_3)^{-1} \\ &= (1 + \sigma[X_1, X_2]/\sigma[X_2, X_3] + \sigma[X_1, X_2]/\sigma[X_1, X_3])^{-1} \text{-----} 2.36.14 \end{aligned}$$

However, Equations 2.36.2 – 2.36.4 in summary state that:

$$\sigma[X_i, X_j] = \lambda_i \lambda_j \sigma^2[T] \text{-----} 2.36.15$$

Isolating the total test true score variance in Equation 2.36.15:

$$\sigma^2[T] = \sigma[X_i, X_j] / \lambda_i \lambda_j \text{-----} 2.36.16$$

Summing Equation 2.36.16 over  $i \neq j$ :

$$\Sigma \Sigma \sigma^2[T] = \Sigma \Sigma (\sigma[X_i, X_j] / \lambda_i \lambda_j) \quad i \neq j = 1, 2, 3 \text{-----} 2.36.17$$

Because the test is divided into three congeneric parts, Equation 2.36.17 can be rewritten as:

$$\begin{aligned} 6(\sigma^2[T]) &= 2(\sigma[X_1, X_2] / \lambda_1 \lambda_2 + \sigma[X_1, X_3] / \lambda_1 \lambda_3 + \sigma[X_2, X_3] / \lambda_2 \lambda_3) \\ &= 2(\sigma[X_1, X_2](1 / \lambda_1 1 / \lambda_2) + \sigma[X_1, X_3](1 / \lambda_1 1 / \lambda_3) + \\ &\quad \sigma[X_2, X_3](1 / \lambda_2 1 / \lambda_3)) \text{-----} 2.36.18 \end{aligned}$$

Substituting the inverse of Equations 2.36.12 – 2.36.14 in Equation 2.36.18:

$$\begin{aligned} 6(\sigma^2[T]) &= 2\{\sigma[X_1, X_2](1 + \sigma[X_2, X_3] / \sigma[X_1, X_3] + \sigma[X_2, X_3] / \sigma[X_1, X_2])(1 + \\ &\quad \sigma[X_1, X_3] / \sigma[X_2, X_3] + \sigma[X_1, X_3] / \sigma[X_1, X_2]) + \\ &\quad \sigma[X_1, X_3](1 + \sigma[X_2, X_3] / \sigma[X_1, X_3] + \sigma[X_2, X_3] / \sigma[X_1, X_2])(1 + \\ &\quad \sigma[X_1, X_2] / \sigma[X_2, X_3] + \sigma[X_1, X_2] / \sigma[X_1, X_3]) + \\ &\quad \sigma[X_2, X_3](1 + \sigma[X_1, X_3] / \sigma[X_2, X_3] + \sigma[X_1, X_3] / \sigma[X_1, X_2])(1 + \\ &\quad \sigma[X_1, X_2] / \sigma[X_2, X_3] + \sigma[X_1, X_2] / \sigma[X_1, X_3])\} \text{-----} 2.36.19 \end{aligned}$$

Multiplying out the round brackets in Equation 2.36.19:

$$\begin{aligned} 6(\sigma^2[T]) &= 2\{\sigma[X_1, X_2](1 + \sigma[X_2, X_3] / \sigma[X_1, X_3] + \sigma[X_1, X_3] / \sigma[X_1, X_2] + \\ &\quad \sigma[X_2, X_3] \sigma[X_1, X_3] / \sigma[X_1, X_3] \sigma[X_2, X_3] + \\ &\quad \sigma[X_2, X_3] \sigma[X_1, X_3] / \sigma[X_1, X_3] \sigma[X_1, X_2] + \\ &\quad \sigma[X_2, X_3] / \sigma[X_1, X_2] + \sigma[X_2, X_3] \sigma[X_1, X_3] / \sigma[X_1, X_2] \sigma[X_2, X_3] + \\ &\quad \sigma[X_2, X_3] \sigma[X_1, X_3] / \sigma[X_1, X_2] \sigma[X_1, X_2]) + \\ &\quad \sigma[X_1, X_3](1 + \sigma[X_1, X_2] / \sigma[X_2, X_3] + \sigma[X_1, X_2] \sigma[X_1, X_3] + \\ &\quad \sigma[X_2, X_3] / \sigma[X_1, X_3] + \sigma[X_2, X_3] \sigma[X_1, X_2] / \sigma[X_2, X_3] \sigma[X_1, X_3] + \\ &\quad \sigma[X_2, X_3] \sigma[X_1, X_2] / \sigma[X_1, X_3] \sigma[X_1, X_3] + \sigma[X_2, X_3] / \sigma[X_1, X_2] + \\ &\quad \sigma[X_2, X_3] \sigma[X_1, X_2] / \sigma[X_1, X_2] \sigma[X_2, X_3] + \\ &\quad \sigma[X_2, X_3] \sigma[X_1, X_2] / \sigma[X_1, X_2] \sigma[X_1, X_3]) + \\ &\quad \sigma[X_2, X_3](1 + \sigma[X_1, X_2] / \sigma[X_2, X_3] + \sigma[X_1, X_2] / \sigma[X_1, X_3] + \end{aligned}$$

$$\begin{aligned} & \sigma[X_1, X_3]/\sigma[X_2, X_3] + \sigma[X_1, X_3]\sigma[X_1, X_2]/\sigma[X_2, X_3]\sigma[X_2, X_3] + \\ & \sigma[X_1, X_3]\sigma[X_1, X_2]/\sigma[X_2, X_3]\sigma[X_1, X_3] + \sigma[X_1, X_3]/\sigma[X_1, X_2] + \\ & \sigma[X_1, X_3]\sigma[X_1, X_2]/\sigma[X_1, X_2]\sigma[X_2, X_3] + \\ & \sigma[X_1, X_3]\sigma[X_1, X_2]/\sigma[X_1, X_2]\sigma[X_1, X_3] \} \text{-----} 2.36.20 \end{aligned}$$

Simplifying Equation 2.36.20:

$$\begin{aligned} 6(\sigma^2[T]) = & 2\{ (2\sigma[X_1, X_2] + 2\sigma[X_1, X_3] + \\ & 2\sigma[X_2, X_3] + \sigma[X_1, X_2]\sigma[X_1, X_3]/\sigma[X_2, X_3] + \\ & \sigma[X_1, X_2]\sigma[X_2, X_3]/\sigma[X_1, X_3] + \sigma[X_1, X_3]\sigma[X_2, X_3]/\sigma[X_1, X_2] + \\ & (2\sigma[X_1, X_2] + 2\sigma[X_1, X_3] + \\ & 2\sigma[X_2, X_3] + \sigma[X_1, X_2]\sigma[X_1, X_3]/\sigma[X_2, X_3] + \\ & \sigma[X_1, X_3]\sigma[X_2, X_3]/\sigma[X_1, X_2] + \sigma[X_1, X_2]\sigma[X_2, X_3]/\sigma[X_1, X_3] + \\ & (2\sigma[X_1, X_2] + 2\sigma[X_1, X_3] + \\ & 2\sigma[X_2, X_3] + \sigma[X_1, X_2]\sigma[X_2, X_3]/\sigma[X_1, X_3] + \\ & \sigma[X_1, X_2]\sigma[X_1, X_3]/\sigma[X_2, X_3] + \\ & \sigma[X_1, X_3]\sigma[X_2, X_3]/\sigma[X_1, X_2] \} \text{-----} 2.36.21 \end{aligned}$$

Simplifying Equation 2.36.21 further:

$$\begin{aligned} 6(\sigma^2[T]) = & 2\{ 6(\sigma[X_1, X_2] + \sigma[X_1, X_3] + \sigma[X_2, X_3]) + \\ & 3(\sigma[X_1, X_2]\sigma[X_2, X_3]/\sigma[X_1, X_3]) + \\ & 3(\sigma[X_2, X_3]\sigma[X_1, X_3]/\sigma[X_1, X_2]) + \\ & 3(\sigma[X_1, X_2]\sigma[X_1, X_3]/\sigma[X_2, X_3]) \} \\ = & 12(\sigma[X_1, X_2] + \sigma[X_1, X_3] + \sigma[X_2, X_3]) + \\ & 6(\sigma[X_1, X_2]\sigma[X_2, X_3]/\sigma[X_1, X_3] + \\ & \sigma[X_2, X_3]\sigma[X_1, X_3]/\sigma[X_1, X_2] + \\ & \sigma[X_1, X_2]\sigma[X_1, X_3]/\sigma[X_2, X_3]) \text{-----} 2.36.22 \end{aligned}$$

Therefore:

$$\begin{aligned} \sigma^2[T] = & 2(\sigma[X_1, X_2] + \sigma[X_1, X_3] + \sigma[X_2, X_3]) + \\ & (\sigma[X_1, X_2]\sigma[X_2, X_3]/\sigma[X_1, X_3] + \\ & \sigma[X_2, X_3]\sigma[X_1, X_3]/\sigma[X_1, X_2] + \\ & \sigma[X_1, X_2]\sigma[X_1, X_3]/\sigma[X_2, X_3]) \text{-----} 2.36.23 \end{aligned}$$

Expressing Equation 2.36.23 in terms of a common denominator:

$$\begin{aligned}
\sigma^2[T] &= (\sigma[X_1, X_2]\sigma[X_1, X_3])/\sigma[X_2, X_3] + \\
&\quad (\sigma[X_1, X_2]\sigma[X_2, X_3])/\sigma[X_1, X_3] + \\
&\quad (\sigma[X_1, X_3]\sigma[X_2, X_3])/\sigma[X_1, X_2] + \\
&\quad 2(\sigma[X_1, X_2] + \sigma[X_1, X_3] + \sigma[X_2, X_3]) \\
&= \{(\sigma^2[X_1, X_2]\sigma^2[X_1, X_3]) + \\
&\quad (\sigma^2[X_1, X_2]\sigma^2[X_2, X_3]) + \\
&\quad (\sigma^2[X_1, X_3]\sigma^2[X_2, X_3])\}/\sigma[X_1, X_2]\sigma[X_1, X_3]\sigma[X_2, X_3] + \\
&\quad 2\{(\sigma[X_1, X_2] + \sigma[X_1, X_3] + \sigma[X_2, X_3])(\sigma[X_1, X_2]\sigma[X_1, X_3]\sigma[X_2, X_3])\} \\
&\quad /(\sigma[X_1, X_2]\sigma[X_1, X_3]\sigma[X_2, X_3]) \text{-----} 2.36.24
\end{aligned}$$

Multiplying out the right-hand side term in Equation 2.36.24:

$$\text{Let } 2\{(\sigma[X_1, X_2] + \sigma[X_1, X_3] + \sigma[X_2, X_3])(\sigma[X_1, X_2]\sigma[X_1, X_3]\sigma[X_2, X_3])\} = G$$

Therefore:

$$\begin{aligned}
G &= 2\sigma^2[X_1, X_2]\sigma[X_1, X_3]\sigma[X_2, X_3] + \\
&\quad 2\sigma[X_1, X_2]\sigma^2[X_1, X_3]\sigma[X_2, X_3] + \\
&\quad 2\sigma[X_1, X_2]\sigma[X_1, X_3]\sigma^2[X_2, X_3] \text{-----} 2.36.25
\end{aligned}$$

Therefore Equation 2.36.24 can be simplified through factorisation:

$$\begin{aligned}
\sigma^2[T] &= \{(\sigma[X_1, X_2]\sigma[X_1, X_3] + \sigma[X_1, X_2]\sigma[X_2, X_3] + \sigma[X_1, X_3]\sigma[X_2, X_3])\} / \\
&\quad (\sigma[X_1, X_2]\sigma[X_1, X_3]\sigma[X_2, X_3]) \text{-----} 2.36.26
\end{aligned}$$

Consequently the reliability coefficient can be written as:

$$\begin{aligned}
\rho_{ttX} &= \sigma^2[T]/\sigma^2[X] \\
&= (\sigma[X_1, X_2]\sigma[X_1, X_3] + \sigma[X_1, X_2]\sigma[X_2, X_3] + \sigma[X_1, X_3]\sigma[X_2, X_3])^2 \\
&\quad /(\sigma[X_1, X_2]\sigma[X_1, X_3]\sigma[X_2, X_3]\sigma^2[X]) \text{-----} 2.36
\end{aligned}$$

Kristof [1974, pp. 491-493] asserts that:

If the parts are homogeneous in content [congeneric], i.e., if their true scores are linearly related and if sample size is large then the method described in this paper will give the precise value of the reliability parameter. If the homogeneity condition is violated then underestimation will typically result. However, the estimate will always be at least as accurate as coefficient  $\alpha$  and Guttman's lower bound  $\lambda_3$  when the same data are used. ... True score variance as given above does not depend on a particular division of the test into three content-



homogenous [congeneric] parts. Any three such parts are admissible and will give the same  $\sigma^2[T]$ . This quantity is not a lower bound. The right-hand side of the  $\sigma^2[T]$  equation is invariant with respect to division of a test into three content-homogenous parts.

Raju [1977] presented a reliability coefficient [coefficient beta] for multiple congeneric measurements equivalent to the coefficient proposed by Gilmer and Feldt [1983]. Raju [1977] assumes an instrument  $S$ , consisting of  $n$  items, partitioned into  $k$  mutually exclusive subtests denoted by  $S_1, \dots, S_i, \dots, S_j, \dots, S_k$  with  $n_1, \dots, n_i, \dots, n_j, \dots, n_k$  items respectively. The proportion of items in subtest  $S_i$  is denoted by  $p_i = n_i/n$ , the subtest scores are denoted by  $X_1, \dots, X_i, \dots, X_j, \dots, X_k$ .

Because  $X = \Sigma X_i$ , and  $X_i = T_i + E_i$ , the observed score variance can be analysed as follows:

$$\begin{aligned}\sigma^2[T] &= \sigma^2[T_1 + T_2 + \dots + T_k] \\ &= \Sigma \sigma^2[T_i] + \Sigma \Sigma \sigma[T_i, T_j] [i \neq j] \\ &= \Sigma \sigma^2[T_i] + \Sigma \Sigma \sigma[X_i, X_j] [i \neq j] \end{aligned} \quad \text{2.37.1}$$

$$\begin{aligned}\text{Since, } \sigma[X_i, X_j] &= \sigma[(T_i + E_i), (T_j + E_j)] \\ &= \sigma[T_i, T_j] + \sigma[T_i, E_j] + \sigma[E_i, T_j] + \sigma[E_i, E_j] \\ &= \sigma[T_i, T_j]\end{aligned}$$

$$\text{Since, } \sigma[T_i, E_j] = \sigma[E_i, T_j] = \sigma[E_i, E_j] = 0$$

Similarly, for subtest  $S_i$  the true score variance can be analysed as follows since the subtest observed score  $X_i$  is itself a linear combination of  $n_i$  observed item scores, each comprising a linear combination of a true score and an error score:

$$\begin{aligned}\sigma^2[T_i] &= \sigma^2[T_{i1} + T_{i2} + \dots + T_{iu} + \dots + T_{iv} + \dots + T_{in_i}] \\ &= \Sigma \sigma^2[T_{iu}] + \Sigma \Sigma \sigma[T_{iu}, T_{iv}] \\ &= \Sigma \sigma^2[T_{iu}] + \Sigma \Sigma \sigma[X_{iu}, X_{iv}] (u \neq v) \end{aligned} \quad \text{2.37.2}$$

Since items are essentially  $\tau$ -equivalent [i.e.  $\tau_u = a + \tau_v$ ], their true score variances are equal and the covariance between any two items is equal to the covariance of any other pair of items. The true score variance of any arbitrary item, furthermore, is equal to the covariance of two arbitrary items.

Therefore:

$$\sigma[X_{iu}, X_{iv}] = \sigma^2[T_{iu}] = \sigma^2[T_{iv}] \text{ for any } u \text{ and } v \quad \text{2.37.3}$$

Let  $g$  represent  $\sigma^2[T_{iu}]$

Equation 2.37.2 can therefore be rewritten as:

$$\begin{aligned}\sigma^2[T_i] &= n_i g + n_i(n_i - 1)g \\ &= n_i g + n_i^2 g - n_i g \\ &= n_i^2 g\end{aligned}\quad \text{-----} \quad 2.37.4$$

Using Equation 2.37.4, Equation 2.27.1 can be rewritten as:

$$\begin{aligned}\sigma^2[T] &= \sum n_i^2 g + \sum \sum \sigma[X_i, X_j] \\ &= g \sum n_i^2 + \sum \sum \sigma[X_i, X_j]\end{aligned}\quad \text{-----} \quad 2.37.5$$

However:

$$\sigma[X_{iu}, X_{iv}] = g$$

Therefore:

$$\sigma[X_i, X_j] = g n_i n_j \quad \text{-----} \quad 2.37.6$$

Therefore:

$$\sum \sum \sigma[X_i, X_j] = g \sum \sum n_i n_j \quad \text{-----} \quad 2.37.7$$

Isolating  $g$  in Equation 2.37.7 results in:

$$g = \sum \sum \sigma[X_i, X_j] / \sum \sum n_i n_j \quad \text{-----} \quad 2.37.8$$

Substituting Equation 2.37.8 into Equation 2.37.5 and multiplying the covariance term with  $\sum \sum n_i n_j / \sum \sum n_i n_j$  results in:

$$\begin{aligned}\sigma^2[T] &= g \sum n_i^2 + \sum \sum \sigma[X_i, X_j] \\ &= \{(\sum n_i^2 \sum \sum \sigma[X_i, X_j]) / \sum \sum n_i n_j\} + \sum \sum \sigma[X_i, X_j] \\ &= \{(\sum n_i^2 \sum \sum \sigma[X_i, X_j]) / \sum \sum n_i n_j\} + (\sum \sum n_i n_j \sum \sum \sigma[X_i, X_j]) / \sum \sum n_i n_j \\ &= \sum \sum \sigma[X_i, X_j] \{(\sum n_i^2 + \sum \sum n_i n_j) / \sum \sum n_i n_j\} \quad \text{-----} \quad 2.37.9\end{aligned}$$

However the total number of items consists of the sum of the number of items in the  $k$  mutually exclusive sub-tests  $S_i$ :

$$n = (n_1 + n_2 + \dots + n_k) \text{-----} 2.37.10$$

Therefore:

$$\begin{aligned} n^2 &= (n_1 + n_2 + \dots + n_k)^2 \\ &= \sum n_i^2 + (i \neq j) \sum \sum n_i n_j \text{-----} 2.37.11 \end{aligned}$$

Therefore when substituting Equation 2.37.11 into Equation 2.37.9 and multiplying the numerator and denominator with  $(1/n^2)$ :

$$\begin{aligned} \sigma^2[T] &= (\sum \sum \sigma[X_i, X_j] n^2) / \sum \sum n_i n_j \\ &= (\sum \sum \sigma[X_i, X_j] n^2 (1/n^2)) / \sum \sum n_i n_j (1/n^2) \\ &= \sum \sum \sigma[X_i, X_j] / \sum \sum n_i n_j / n^2 \text{-----} 2.37.12 \end{aligned}$$

Let  $p_i$  denote the proportion of items in  $S_i$ . Therefore:

$$p_i = n_i/n \text{ and } p_j = n_j/n \text{-----} 2.37.13$$

Therefore:

$$p_i p_j = n_i n_j / n^2 \text{-----} 2.37.14$$

Substituting Equation 2.37.14 into Equation 2.37.12:

$$\sigma^2[T] = \sum \sum \sigma[X_i, X_j] / \sum \sum p_i p_j \text{-----} 2.37.15$$

Therefore:

$$\sigma^2[T] / \sigma^2[X] = \sum \sum \sigma[X_i, X_j] / (\sum \sum p_i p_j \sigma^2[X]) \text{-----} 2.37.16$$

However:

$$\begin{aligned} \sigma^2[X] &= \sigma^2[X_1 + X_2 + \dots + X_k] \\ &= \sum \sigma^2[X_i] + \sum \sum \sigma[X_i, X_j] \text{-----} 2.37.17 \end{aligned}$$

Isolating the covariance term in Equation 2.37.17:

$$\Sigma \Sigma \sigma[X_i, X_j] = \sigma^2[X] - \Sigma \sigma^2[X_i] \text{-----} 2.37.18$$

Substituting Equation 2.37.18 into Equation 2.37.16:

$$\sigma^2[T]/\sigma^2[X] = (\sigma^2[X] - \Sigma \sigma^2[X_i]) / (\Sigma \Sigma p_i p_j \sigma^2[X]) \text{-----} 2.37.19$$

However:

$$\Sigma p_i = 1$$

Therefore:

$$[\Sigma p_i]^2 = 1^2$$

Therefore:

$$(p_1 + p_2 + \dots + p_k)^2 = 1$$

Therefore:

$$\Sigma p_i^2 + \Sigma \Sigma p_i p_j = 1$$

Therefore:

$$\Sigma \Sigma p_i p_j = 1 - \Sigma p_i^2 \text{-----} 2.37.20$$

Substituting Equation 2.37.20 into Equation 2.37.19:

$$\sigma^2[T]/\sigma^2[X] = (\sigma^2[X] - \Sigma \sigma^2[X_i]) / \{(1 - \Sigma p_i^2) \sigma^2[X]\}$$

Therefore we can write Equation 2.37.16:

$$\begin{aligned} \beta_k &= \Sigma \Sigma \sigma[X_i, X_j] / (\sigma^2[X] \Sigma \Sigma p_i p_j) \\ &= (\sigma^2[X] - \Sigma \sigma^2[X_i]) / (\sigma^2[X] \Sigma \Sigma p_i p_j) \\ &= (\sigma^2[X] - \Sigma \sigma^2[X_i]) / \{\sigma^2[X] (1 - \Sigma p_i^2)\} \text{-----} 2.37 \end{aligned}$$

The Raju beta coefficient [ $\beta_k$ ] like the Gilmer and Feldt coefficient, will equal the Cronbach  $\alpha$  if the  $k$  components are of equal length so that all  $p_i$  equal  $1/k$ . Otherwise  $\beta_k$  will exceed  $\alpha$  for all values of  $k$  [Raju, 1977]. For  $k = 2$  congeneric measurements, Equation 2.37 reduces to the formula derived by Raju [1970] shown as Equation 2.26. For  $k = 3$  congeneric measurements, Equation 2.37, reduces to a formula shown as Equation 2.38 which represents an alternative to the Kristof [1974] reliability estimate [Equation 2.36].

$$\beta_3 = (\sigma[X_1,X_2]+\sigma[X_1,X_3]+\sigma[X_2,X_3])/\{\sigma^2[X](p_1p_2+p_1p_3+p_2p_3)\}----- 2.38$$

2.5.1.1.3 Parallel Forms Method

The third practical method for the estimation of the reliability of a set of observed scores with parallel measurements requires the construction of two forms of the same test consisting of items that are different but provide [approximately] parallel measurements when applied [Lord & Novick, 1968]. The term coefficient of equivalence is normally used to refer to the correlation between the two [approximately] parallel test forms. The coefficient of equivalence would reflect error variance produced by the imprecision of the measurement and the lack of parallelism between the [approximately] parallel test forms. Successive administrations of parallel test forms, however, permit the effects of fatigue, practice and true score changes to operate. In reality the application of parallel forms thus would produce a coefficient of stability and equivalence which reflects error due to imprecision, variation between forms and variation of testees' ability between applications. The method, furthermore, has the practical limitation that relatively few tests are published with [approximately] parallel test forms. In addition, the development of [approximately] parallel test forms, represents a sufficiently formidable task to dissuade the majority of test users from attempting it themselves.

A summary of the various empirical internal-consistency reliability estimators is shown in tabular form in Table 2.2.

Table 2.2. Summary of internal-consistency reliability estimators

MEASUREMENT MODEL	NAME	FORMULA
Two classical halves	Spearman-Brown	$\rho_{ttX}=2\rho[X_{11},X_{12}]/1+\rho[X_{11},X_{12}]$
Multiple classical components	Generalised Spearman-Brown	$\rho_{ttX}=kp[X_i,X_j]/1 + [k-1]\rho[X_i,X_j]$
Two essentially $\tau$ -equivalent halves	Cronbach $\alpha$ /Guttman $\lambda_4$	$\rho_{ttX}=2\{1-(\sigma^2[X_1]+\sigma^2[X_2])/\sigma^2[X]\}$
Two essentially $\tau$ -equivalent halves	Flanagan	$\rho_{ttX}=4\sigma[X_1,X_2]/\sigma^2[X]$

Two essentially $\tau$ -equivalent halves	Rulon	$\rho_{ttX} = 1 \cdot \sigma^2[X_1 \cdot X_2] / \sigma^2[X]$
Two classical halves	Rulon	$\rho_{ttX} = 1 \cdot \{\sigma^2[X_1 \cdot X_2] / \sigma^2[X]\}$
Two congeneric halves	Raju	$\rho_{ttX} = \sigma[X_1, X_2] / \lambda_1 \lambda_2 \sigma^2[X]$
Two congeneric halves	Feldt	$\rho_{ttX} = 4\sigma[X_1, X_2] / \{\sigma^2[X] - \{(\sigma^2[X_1] + \sigma^2[X_2]) / \sigma[X]^2\}\}$
Essentially $\tau$ -equivalent components	Cronbach $\alpha$	$\alpha = [k / (k-1)] [1 - \Sigma \sigma^2[X_i] / \sigma^2[X]]$
Essentially $\tau$ -equivalent components	Kuder-Richardson-20	$\rho_{ttX} = [k / (k-1)] [1 - \Sigma P_i Q_i / \sigma^2[X]]$
Essentially $\tau$ -equivalent components	Kuder-Richardson-21	$\rho_{ttX} = [k / (k-1)] [1 - k\mu(P)\mu(Q) / \sigma^2[X]]$
Essentially $\tau$ -equivalent components	Guttman $\lambda_2$	$\lambda_2 = [\Sigma \Sigma \sigma[X_i, X_j] + \{k / (k-1) \Sigma \Sigma \sigma^2[X_i, X_j]\}^{1/2}] / \sigma^2[X]$
Multiple congeneric	Gilmer & Feldt	$\rho_{ttX} = \{1 / (1 - \Sigma \lambda_i^2)\} \{1 - \Sigma \sigma^2[X_i] / \sigma^2[X]\}$
Multiple congeneric	Kristof	$\rho_{ttX} = \{\sigma[X_1, X_2]\sigma[X_1, X_3] + \sigma[X_1, X_2]\sigma[X_2, X_3] + \sigma[X_1, X_3]\sigma[X_2, X_3]\}^2 / \sigma[X_1, X_2]\sigma[X_1, X_3]\sigma[X_2, X_3]\sigma^2[X]$
Multiple congeneric	Raju $\beta_k$	$\beta_k = \{\sigma^2[X] - \Sigma \sigma^2[X_i]\} / \{\sigma^2[X](1 - \Sigma p_i^2)\}$

## 2.5.2 Validity

Validity is a concept of fundamental importance to psychological measurement. Validity is, however, also a concept of considerable complexity. Despite this, various naïve/limited, if not erroneous, conceptions of validity have developed to detrimentally effect the debate on personnel decisions based on psychological measurements [Binning & Barrett, 1989; Schmitt & Landy, 1993]. This tendency seems to apply especially to human resource practitioners in South Africa.

The theoretical conceptualisation of validity has evolved over the years [Landy, 1986; Messick, 1989, Schmitt & Landy, 1993], shaped by the combined influence of changing legal, social, economic and technical considerations on the quality of psychological measurement. Since at least the early 1950's, test validity has been broken into three distinct types, one of which comprises two subtypes [Messick, 1989]. These are the familiar trinity of content validity, criterion related validity [subsuming predictive and concurrent validity] and construct validity [Ellis & Blustein, 1991; Landy, 1986; Messick, 1989, Schmitt & Landy, 1993; Schuler & Guldin, 1991]. The taxonomy itself is not fundamentally flawed [Landy, 1986] in as far as it suggests that different inferences can be made from test scores. The linkage of these validity concepts to specific aims of testing by the American Psychological Association in their technical recommendations on psychological testing [APA, 1954; APA, 1966; APA, 1974], in conjunction with Title VII litigation case law [Landy, 1986], did however create the erroneous trinitarian notion that only a single validation type or strategy need to be considered in any given



situation. Although each subsequent revised edition of the Standards for educational and psychological tests/testing did to a certain, albeit diminishing, degree perpetuate this erroneous distinction between validity types, they did, however, gradually move towards a unificationist perspective on the validity concept [Messick, 1989].

A central thesis developed in chapter 1 is that measurements are obtained by reason of the need for quantitative information on constructs for the purpose of decision-making. The relevance of information on a construct for a particular decision is, in the final analysis, based on logical and ethical considerations rather than empirical psychometric evidence. The question, which constructs to obtain information on, therefore, is not a psychometric question.

Due to the abstract nature of psychological constructs, information on their states or levels can be inferred only from their observable effects. The question consequently arises to what extent the operationalisation of the relevant construct in terms of observable denotations was successful. The pivotal question is to what extent the measurement procedure produces a pure [i.e. uncontaminated by irrelevant influences] and comprehensive [i.e. complete coverage of all facets of the construct as constitutively defined] quantitative description of the [multi-dimensional] construct regarded as relevant for the decision at hand. An alternative, more prevalent formulation of the above interpretation would be to define validity as the degree to which an instrument measures the construct which it claims or intends to measure [Nunnally, 1978]. Validity thus refers to the degree to which variance in a set of measures is due to variance in the underlying construct of interest [Guion, 1965]. The validity of a set of measurements refers to the formal consideration of the preceding question. If all personnel decisions could be made in terms of quantitative descriptions of constructs measurable at the time of the decision, the foregoing basic conceptualisation of validity would probably have sufficed. This is, however, not the case. For an important class of personnel decisions information on the construct deemed relevant to the decision cannot be obtained prior to the decision.

To provide for this eventuality, while still remaining relevant to the earlier discussion, validity can more generally be interpreted as the extent to which the inferences made from the test scores are in fact warranted; the extent to which the interpretation [i.e. meaning] assigned to test scores is justified/supported. Strictly speaking, therefore, what is being validated is not the measuring instrument, nor the measures obtained from the instrument, but rather the inferences made from the measures<sup>1</sup>. Messick [1989, p. 13], in his monumental and definitive treatment of the validity concept, states:

---

<sup>1</sup> The same principle also applies for any other context in which the validity concept is used; research designs, for example, can be designated internally valid to the extent that they permit causal inferences of the form  $Y=f[X]$ .

Validity is an integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment. .... Broadly speaking, then, validity is an inductive summary of both the existing evidence for and the potential consequences of score interpretation and use. Hence what is to be validated is not the device as such but the inferences derived from test scores or other indicators - inferences about score meaning or interpretation and about the implications for action that the interpretation entails.

Validation, in turn, refers to the process of accumulating and cementing empirical evidence and logical thought in a credible argument in defense of the inferences made from the observed scores. Kane [1992, p. 527] expresses his views on an argument-based approach to validation as follows:

A test-score interpretation always involves an interpretive argument ..... To validate a test-score interpretation is to support the plausibility of the corresponding interpretive argument with appropriate evidence. The argument-based approach to validation adopts the interpretive argument as the framework for collecting and presenting validity evidence and seeks to provide convincing evidence for its inferences and assumptions, especially its most questionable assumptions.

The type of validity evidence required to justify the inferences made from test scores depends on the nature of the interpretive argument. This should, however, not be construed to mean that a single validity analysis strategy [Lawsche, 1985] should be chosen from the trilogy of content, construct and criterion-related validities [Ellis & Blustein, 1991; Landy, 1986]. The different validity analysis strategies are not alternatives but rather form supplementary facets of a single unitary validity concept [Binning & Barrett, 1992; Ellis & Blustein, 1991; Guion, 1991; Messick, 1989; Schmitt & Landy, 1993]. Schmitt and Landy [1993, p. 286] clearly affirm the foregoing position by stating:

Marshaling evidence of validity is now seen as a process of theory development and testing [Binning & Barrett, 1989; Landy, 1986]. We must develop and articulate theories of job performance and define logically the constructs that are central to these theories. We must establish a 'nomological network' that relates constructs important in the job performance domain to the constructs we choose to identify qualified job applicants. This requires evidence that the measures we use to operationalize constructs in the predictor and performance domains possess a logical relationship to these constructs and empirically consistent relationships to other measures of the construct.

The recognition of the inappropriateness of using the trinitarian conceptualisation of validity in a way that resembles "stamp collecting" [Landy, 1986], in essence rests on the realisation that personnel selection decisions are not based on a single inference but rather an integrated network of inferences or a multi-step argument. In personnel selection decisions, job performance forms the basis [i.e., the criterion] on which applicants should be evaluated so as to determine their assignment to an

appropriate treatment [Cronbach, 1965]. Information on actual job performance can, however, never be available at the time of the selection decision. This seemingly innocent, but too often ignored, dilemma provides the impetus for the development of the aforementioned integrated network of inferences or multi-step argument. Even though it is logically impossible to obtain direct information on the performance construct [i.e., the final/ultimate criterion] at the time of the selection decision, it can nonetheless be predicted at the time of the selection decision if:

- variance in the performance construct can be explained in terms of one or more predictors;
- the nature of the relationship between these predictors and the performance construct has been made explicit; and
- predictor information can be obtained prior to the selection decision.

The point of departure for the development of any personnel selection procedure is the performance construct [Binning & Barrett, 1989]. Job analysis provides the necessary information on the job content and context to constitutively define the performance construct. Two qualitatively different arguments exist in terms of which predictors can be derived from the conceptualisation of job performance. Binning and Barrett [1989] seem to be of the same opinion. Although they initially refer to “three routes from the performance domain to predictor development” they later propose that “the construct-related and content-related approaches represent the two fundamental predictor sampling strategies” [Binning & Barrett, 1989, p. 483]. This position, previously described in chapter 1, correlates with a distinction Wernimont and Campbell [1968] makes between predictors as signs and predictors as samples. Since these two different approaches to predictor development rely on quantitatively different arguments they have to be justified in terms of different types of validity evidence.

A construct-related approach to predictor development utilises the conceptualisation of the performance construct in conjunction with theory and logic to develop through theorising a complex performance hypothesis [a tentative nomological network] as a tentative performance theory. If the complex performance hypothesis would in fact be valid, it would in principle be possible to estimate job performance as a substitute for actual job performance [provided the nature of the relationship between the performance construct and its person-centered determinants would also be known]. Personnel selection procedures are thus possible in terms of this approach only if they are based on a valid substantive performance hypothesis. The efficiency of such procedures would in turn depend on to the extent to which the underlying performance hypothesis reflects the full complexity of the forces shaping job performance [both in terms of the nature of the determinants and the way they combine]. To establish the validity of the performance hypothesis, an operational hypothesis is deductively derived from the substantive performance hypothesis by operationally defining the performance construct and the explanatory psychological constructs. The operational definition of the performance



construct constitutes a premise in the aforementioned deductive argument, as do the operational definitions of the explanatory psychological constructs. The validity of the deductive argument depends on the validity of these premises [Copi & Cohen, 1990; Mouton & Marais, 1985]. In a valid deductive argument the premises provide conclusive grounds for the truth of the conclusion [Copi & Cohen, 1990]. The justifiability of the claim that the operational performance hypothesis constitutes a valid testable representation of the theoretical performance hypothesis thus depends on the construct validity of the operational measures of the performance construct and the explanatory psychological determinants. Should empirical confirmation for the operational performance hypothesis be found [assuming that the aforementioned deductive argument was in fact valid], the substantive performance hypothesis may be considered corroborated since it has survived an opportunity to be refuted [Popper, 1972]. The validity of the substantive performance hypothesis, in conjunction with evidence on the construct validity of the operational measures of the explanatory psychological constructs, provides justification for the claim that job performance can be inferred from an array of operational predictor measures developed through a construct-related approach. Binning and Barrett [1989] consequently propose five inferences or hypotheses to be central to the validation of a personnel selection procedure, namely:

- the [multidimensional] performance/criterion construct is related to [and thus could in principle be inferred from] an array of systematically interrelated [possibly also multidimensional] predictor or person-centered constructs [linkage 1]; and
- a[n] [multidimensional] operational criterion measure provides a pure and comprehensive empirical grasp on the [multidimensional] performance construct [linkage 2] so as to practically enable the inference of the latter from the former; and
- an array of operational predictor measures provide empirical grasps on the array of corresponding predictor constructs [linkage 3] so as to practically enable the inference of the latter from the former; and consequently
- the operationalised performance/criterion construct is related to [and thus could be inferred from] an array of systematically interrelated operationalised predictor or person-centered constructs [linkage 4], and consequently
- the [multidimensional] performance construct is related to [and thus could be inferred from] an array of systematically interrelated operational predictor or person-centered measures [linkage 5].

These five pivotal linkages in the interpretive argument [Kane, 1992] or nomological network [Schmitt & Landy, 1993] underlying selection instrument score interpretation in the context of personnel selection are depicted in Figure 2.1. In personnel selection research the inference of prime importance is linkage 5. The critical inference that needs to be justified is therefore the linkage [linkage 5] between the predictor battery and the theoretical criterion/performance construct [Binning & Barrett, 1989;

Schmitt & Landy, 1993]]. The claim that actual job performance can be inferred from information obtained from selection techniques, needs to be substantiated.

The criterion-related validity coefficient examines the validity of the hypothesised linkage 4. The criterion-related validity coefficient, however, only constitutes a necessary, though not sufficient, element in the evidence required to justify performance inferences from information obtained on the predictor battery and thus to vindicate a selection procedure should it be challenged in terms of either efficiency or equity [Milkovich & Boudreau, 1991]. Evidence must also be lead in support of the hypothesised linkages 2 and 3. The construct validity of the operational predictor and performance measures must therefore also be established.

A content-related approach to predictor development also utilises the conceptualisation of the performance construct as the basic point of departure, but now with the purpose of selecting a representative sample of tasks or demands from the performance domain. A content-related approach to predictor development requires the development of [low or high fidelity] simulations of the tasks or demands comprising the performance domain. Inferences regarding future job performance may justifiably be made from such predictors if sufficient evidence exists that the simulations comprehensively cover the job performance domain.

Successful performance on the simulation implies successful performance on the job if the demand set put to the applicant is equivalent to the demand set put to the job incumbent. Therefore, under the content-related approach to predictor development, the content validity of predictors needs to be established through empirical and logical analysis. With reference to Figure 2.1, p. 101 the inference of prime importance is, therefore, still linkage 5. Confidence in linkage 5 can or should, however, be bolstered by empirically showing that the operational performance measure correlates with the predictor [i.e., by establishing the criterion-related validity of the predictor or linkage 4]. This in turn, however, necessitates empirical evidence on the construct validity of the operational performance/criterion measure [linkage 2] and the content validity of the predictor measure.

The construct-related and content-related approaches to predictor development clearly differ in terms of the argument through which they derive selection assessment techniques. It would consequently be tempting to infer that substantive differences should also exist between construct-related and content-related assessment techniques. This is, however, not the case. Whether the selection technique is a traditional, construct orientated, psychological test, a low fidelity simulation like a behavioural event interview, or a high fidelity simulation like virtual reality flying, the essential structure of the measurement process is the same. A sample of observable behaviour is elicited through a sample of standardised stimuli under standardised conditions. The elicited behaviour is subsequently observed, recorded and quantified through a standardised procedure.

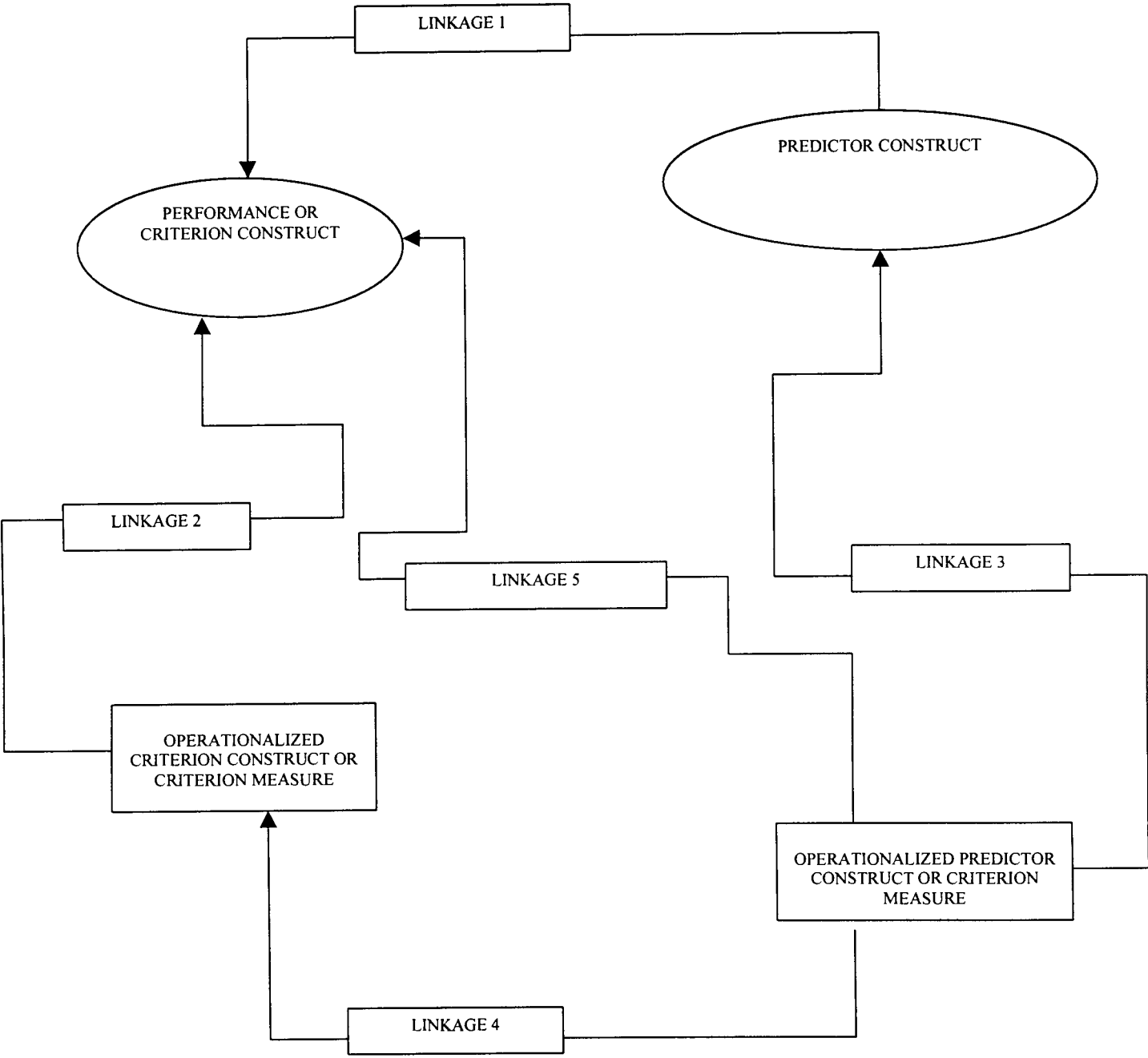


Figure 2.1. The five pivotal linkages in the interpretive argument underlying selection instrument score interpretation in the context of personnel selection [Binning & Barrett, 1989, p. 480; Schmitt & Landy, 1993, p. 287].



The nature and appearance of the [test] stimuli might be different. The logic and processes through which the stimuli were constructed are necessarily different.

Under a construct-related approach the stimulus sample is constructed so as to reflect, through the testee's behavioural reaction to it, a specific construct presumed to be relevant to job performance. Under a content-related approach the stimulus sample is constructed so as to represent the salient job tasks or demands and thereby to indicate, through the testee's behavioural reaction to it, the applicants suitability for the job. For a specific job, however, and assuming both types of selection techniques to have been equally successfully [though differently] justified, the stimulus sample constructed under the content-related approach would evoke the same psychological constructs than the stimulus sample constructed under the construct-related approach would. Irrespective of the approach used to develop a selection technique, if a selection technique is to have any value at all it must cover the same psychological constructs that underlie job performance. Under a content-related approach the nature of the underlying psychological constructs would/need not be known while the explication of the nature of the psychological constructs underlying job performance forms the pivot around which the construct-related approach revolves.

This line of reasoning has additional important implications. Selection techniques developed through a content-related approach are often perceived to be immune to threats of measurement bias and decisions based on information obtained from such selection techniques are often seen not to be susceptible to fairness problems. This seems to be especially true if the selection technique is labeled as a competency-based approach to assessment. Clearly there exists no psychometric justification for such beliefs if the essential structure of the measurement process remains the same. In principle the concepts of bias and fairness apply to assessment techniques derived through a content-related approach as much as they apply to assessment techniques derived through a construct-related approach.

From the preceding discussion it is readily apparent that criterion-related validity represents a research strategy for empirically assessing the quality of both the construct-related and content-related approaches to predictor development. The criterion-related research strategy, furthermore, constitutes an indispensable element under both the construct-related and content-related approaches because of the need to map the operational predictor domain on the operational performance domain. The search for predictors was initiated because of the unavailability of actual job performance information at the time of the selection decision and the subsequent need to estimate job performance. All selection assessments, therefore, need to be interpreted in terms of expected criterion performance [i.e.,  $E[Y | X]$ ] irrespective of the approach through which the assessment technique was developed. This can only be achieved if the nature of the relationship between the operationalised performance and predictor domains are explicitly examined and described.

The criterion-related validity coefficient of a set of measures  $X_j$  obtained from the application of test form  $j$  to  $n$  testees with respect to a criterion  $Y_j$  is defined as the value of correlation coefficient shown by Equation 2.39.

$$\rho[X,Y] = \sigma[X,Y]/\sigma[X]\sigma[Y] \text{-----} 2.39$$

The criterion-related validity coefficient is interpreted in terms of the coefficient of determination  $\rho^2[X,Y]$ , shown as Equation 2.40.

Assume the regression of  $Y$  on  $X$  to be linear. Consequently the expected criterion performance conditional on  $X$  can be expressed as follows:

$$E[Y | X] = E[Y] + \rho[X,Y](\sigma[Y]/\sigma[X])(X-E[X]) \text{-----} 2.40.1$$

Subtracting  $E[Y]$  on both sides of Equation 2.40.1:

$$E[Y | X] - E[Y] = \rho[X,Y](\sigma[Y]/\sigma[X])(X-E[X]) \text{-----} 2.40.2$$

Squaring Equation 2.40.2:

$$(E[Y | X] - E[Y])^2 = \rho^2[X,Y](\sigma^2[Y]/\sigma^2[X])(X-E[X])^2 \text{-----} 2.40.4$$

Summing over the population of observations:

$$\Sigma(E[Y | X] - E[Y])^2 = \rho^2[X,Y](\sigma^2[Y]/\sigma^2[X])\Sigma(X-E[X])^2 \text{-----} 2.40.5$$

Isolating  $\rho^2[X,Y]$  in Equation 2.40.5:

$$\rho^2[X,Y] = \{\Sigma(E[Y | X] - E[Y])^2/\Sigma(X-E[X])^2\}\{\sigma^2[X]/\sigma^2[Y]\} \text{-----} 2.40.6$$

Simplifying Equation 2.40.6:

$$\begin{aligned} \rho^2[X,Y] &= \{\Sigma(E[Y | X] - E[Y])^2/\Sigma(X-E[X])^2\}\{\Sigma(X-E[X])^2/n\}\{\Sigma(Y-E[Y])^2/n\} \\ &= \{\Sigma(E[Y | X] - E[Y])^2/\Sigma(Y-E[Y])^2/n\} \\ &= \sigma^2[E[Y | X]]/\sigma^2[Y] \text{-----} 2.40 \end{aligned}$$

The coefficient of determination thus equals the ratio of the variance in the criterion that can be explained in terms of the regression of  $Y$  on  $X$  relative to the criterion variance.

Although validity and reliability are separate, distinguishable concepts, they are nonetheless conceptually linked. As shown below [Equation 2.41] the reliability of a test could simply be interpreted as its validity with respect to a parallel test [Lord & Novick, 1968].

Using the definition formula for the Pearson correlation coefficient as point of departure:

$$\rho[X,Y] = \sigma[X,Y]/\sigma[X]\sigma[Y] \text{-----} 2.41.1$$

Given Equation 2.1, the covariance term in Equation 2.41.1 can be analysed as follows:

$$\begin{aligned} \sigma[X,Y] &= \sigma[(T_X + E_X),(T_Y + E_Y)] \\ &= \sigma[T_X,T_Y] + \sigma[T_X,E_Y] + \sigma[E_X,T_Y] + \sigma[E_X,E_Y] \\ &= \sigma[T_X,T_Y] \text{-----} 2.41.2 \end{aligned}$$

Equation 2.41.1 can therefore be rewritten as:

$$\rho[X,Y] = \sigma[T_X,T_Y]/\sigma[X]\sigma[Y] \text{-----} 2.41.3$$

If X and Y are classically parallel measurements, then:

$$\sigma[T_X,T_Y] = \sigma^2[T] \text{ and } \sigma[X] = \sigma[Y]$$

Consequently, Equation 2.41.3 can be rewritten as:

$$\begin{aligned} \rho[X,Y] &= \sigma^2[T]/\sigma^2[X] \\ &= \rho_{ttX} \text{-----} 2.41 \end{aligned}$$

## 2.6 SUMMARY

The purpose of chapter 2 was to systematically unfold a formal explication of the classical measurement theory so as to provide a theoretical model in terms of which the derivation of correction formula can be formally demonstrated and comprehensively discussed. It was argued that the core of a selection procedure is the performance theory/hypothesis on which the selection procedure is based. The performance hypothesis is a set of interrelated constructs, their definitions and propositions about the presumed interrelationship between them. The chapter discussed the nature and role of constructs in psychological theory.

Justification of a selection procedure requires empirical confirmation of the validity of propositions contained in the performance hypothesis. To empirically investigate the validity of such a performance hypothesis, however, requires information on the constructs comprising the hypothesis. Due to the abstract character of constructs, information on their state or level can be estimated only from their effects. Thus measurement of constructs by necessity is of an indirect, inferential nature, through the observation of indicants of the construct assumed or demonstrated to be related to the property. The chapter examined and discussed the nature of psychological measurement.

The acceptance of operationalism as a solution to the measurement problem posed by the abstract nature of psychological constructs rests on the supposition that inter- and intra-individual variance in the observed indicant can be explained solely in terms of differences in the underlying construct of interest. Although this represents a practically unattainable ideal, it nevertheless defines the objective of perfectly controlled measurement in which all extraneous variables are controlled. Control is pursued through two processes aimed at either removing the irrelevant variables or keeping the irrelevant variable constant; the effect of both being that the variables no longer produce variance in observed scores. The two processes in question are standardisation and test construction/item analysis. Perfect control, however, is never achieved. The question, consequently, arises to what extent these processes did succeed in controlling extraneous variance. Reliability and validity constitute two evaluatory standards employed to answer this question. The chapter defined reliability and subsequently presented the derivation of a number of reliability coefficients. In addition the chapter defined validity and presented the criterion-related validity coefficient and its interpretation.

ξψζξψζξψζξψζξψζξψζξψζξψζξψ

## CHAPTER 3

### SEPARATE AND COMBINED CORRECTIONS TO THE VALIDITY COEFFICIENT FOR THE ATTENUATING EFFECT OF THE UNRELIABILITY OF MEASUREMENTS AND RESTRICTION OF RANGE

The purpose of the following chapter is firstly to systematically motivate the need for corrections to the validity coefficient for the attenuating effect of the unreliability of the measurements and/or restriction of range. Subsequently the derivation of the various currently available correction formula are discussed and the conditions under which their application would be appropriate.

#### 3.1. THE NEED FOR CORRECTIONS TO THE VALIDITY COEFFICIENT FOR THE ATTENUATING EFFECT OF THE UNRELIABILITY OF MEASUREMENTS AND RESTRICTION OF RANGE

The process of validating a selection procedure in essence refers to a simulation of the selection procedure so as to generate the empirical evidence needed to construct a convincing argument in terms of which the appropriateness of the inferences, decisions and actions based on test scores in the context of actual application can be justified [Guion, 1991; Kane, 1992; Messick, 1989]. Inferences, decisions and actions based on test scores are the outcomes of an implicit or explicit interpretive argument [Kane, 1992]. At the core of the argument lies an array of hypothesis constituting a tentative theory of work performance. Consequently, what needs to be validated for the applicant population is the hypothesis that variance in ultimate criterion performance [Blum & Naylor, 1968] can be inferred from scores on the selection tests [Guion, 1991]. The credibility of assertions on the validity of the performance hypothesis in turn hinges on the unassailability of the research methodology in terms of which these hypothesis are empirically investigated. Cook, Campbell and Peracchio [1991] identify four types of validity that, to the extent that they are threatened, render the research argument vulnerable to questioning and consequently jeopardise the credibility of assertions on the validity of the performance hypothesis. The degree of confidence with which variance in the dependent/criterion construct can be explained in terms of one or more hypothesised independent constructs is contingent on the extent to which the research methodology/research design controls threats to the following four kinds of validity [Cook, Campbell & Peracchio, 1992]:

- statistical conclusion validity;
- construct validity;
- internal validity; and
- external validity.



Statistical conclusion validity, a neologism created with some self-confessed trepidation by Cook, Campbell and Peracchio [1991], refers to the validity of conclusions/decisions on the attributability of the observed sample results [i.e. covariance or difference] to chance. Construct validity refers to the validity with which the measurement operations designed to provide an empirical grasp on the dependent and independent constructs of interest, do in fact measure these constructs as constitutively defined. Internal validity refers to the validity of the inference that variance in the dependent/criterion variable would be increased/decreased by an increase/decrease in the variance of the hypothesised independent variable. Internal validity is consequently concerned with the question whether the observed covariance between X-as measured [or X-as manipulated] and Y-as measured can logically be explained in terms of a cause-effect relationship. External validity concerns the generalisability of research findings across times, conditions and individuals [Campbell & Stanley, 1963; Cook, Campbell & Peracchio, 1991]. In an applied context external validity refers to the transportability of the research findings to the area of application, that is to the [problem] area that motivated the research in the first place.

The primary objective driving validation research is the accumulation of [credible] empirical evidence in terms of which the appropriateness of the inferences, decisions and actions based on test scores in the context of actual application can be justified [Guion, 1991; Kane, 1992; Messick, 1989]. The focus of validation research thus falls on the actual utilisation of the selection procedure. The times, conditions and individuals to which the validation research findings should be transported are consequently those defining the area of actual application. Should the conditions and individuals characterising the validation study in one or more respects not reflect those that would apply at the time of actual application and, should the validation results be sensitive to the features that differ across the areas of analysis and application, the transportability of the research evidence becomes questionable [i.e. the external validity of the validation design would be low]. Consequently the credibility of the argument in terms of which the validation researcher aims to justify the use of the selection procedure, is reduced due to the questionable relevance of the validation evidence mustered. Previously the following specific threats to the external validity of a validation research design were distinguished [Campbell, 1991; Lord & Novick, 1968; Tabachnick & Fidell, 1989;]:

- the extent to which the actual or operationalised criterion contains random measurement error;
- the extent to which the actual or operationalised criterion is systematically biased; i.e. the extent to which the actual criterion is deficient and/or contaminated [Blum & Naylor, 1968];
- the extent to which the validation sample is an unrepresentative, biased, sample from the applicant population in terms of homogeneity and specific attributes [e.g. motivation, knowledge/experience]; and



- the extent to which the sample size and the sample size to number of predictors ratio allow capitalisation on chance and thus overfitting of data.

At the core of the to be constructed argument lies an array of hypotheses constituting a tentative theory of work performance. Thus, what needs to be validated for the applicant population, is the hypothesis that variance in ultimate [i.e. no contamination and no deficiency] criterion performance [Blum & Naylor, 1968] can be explained in terms of variance in scores obtained from selection tests [Guion, 1991] representing construct valid operationalisations of the independent constructs suggested by the hypothesised theory of work performance. The traditional statistical parameter in terms of which the array of operationalised research hypotheses would normally be translated into statistical hypotheses, is the Pearson correlation coefficient. Validation research thus essentially represents an attempt to obtain an unbiased estimate of the parametric validity coefficient from the sample statistic derived from the validation study. The foregoing argument thus translates to the notion that should factors exist that bias the sample estimate of the parameter it would render the argument developed in defense of the selection procedure vulnerable to attack except if such factors would be formally incorporated in the psychometric defense of the selection procedure.

Aside from criterion deficiency and sample unrepresentativeness [i.e. the initial, unrestricted applicant group represents an unrepresentative sample from the applicant population], it is possible to statistically correct for all the other aforementioned threats to external validity of a validation design, provided the required information is in fact available [which in principle is possible, although unlikely]. The need for these corrections to the validity coefficient are motivated in terms of threats to the external validity of the research design characterising a validation study. The obtained validity coefficient cannot, without appropriate corrections, be considered an unbiased estimate of the parametric validity coefficient of interest. The parameter of interest is the correlation coefficient obtained when the sample weights derived from a representative sample are applied to the applicant population and the weighted composite score would be correlated with error-free measurements [i.e. no intermediate criterion contamination [random or systematic] or deficiency] of the [effectively ultimate] criterion for the applicant population [Campbell, 1991]. Guion [1991, p. 351] expresses the importance of considering the transportability of the sample validation evidence to the actual area of application as follows:

Use of a sample statistic to estimate the population parameter is ordinarily accepted if the measurements are reliable, the sample large, and the sample of that population unbiased. These conditions are not routinely found in validation research. Questions of research design are, ultimately, questions of how and whether methods of data collection and data analysis permit inferences from the sample at hand about a larger applicant population.

The validity coefficient should, logically therefore, be calculated after partialling out criterion variance attributable to irrelevant contaminating factors [assuming they are known and can be measured] and subsequently corrected for restriction of range, criterion unreliability and the fitting of error by differential predictor weights, before being transported to the area of application. The current study will, however, only consider the separate and combined effect of corrections for restriction of range and criterion unreliability [i.e. random criterion contamination].

The ever present attenuating effect of the unreliability of the criterion measurements and, the practically always present, inflationary effect of capitalisation on chance and thus overfitting of data occurs irrespective of the underlying research/validation design. Consequently corrections for the attenuating effect of the unreliability of the criterion and shrinkage corrections will almost always apply. The necessity and feasibility of correcting the obtained validity coefficient for the attenuating effect of restriction of range, however, depends on the research design employed.

Tiffen [1946] apparently [Barrett, Philips & Alexander, 1981] was the first to distinguish between concurrent and predictive validation designs. Both Guion and Cranny [1982] and Sussmann and Robertson [1986], however, argue that the simple dichotomous distinction proposed by Tiffen [1946] does not provide a satisfactorily comprehensive coverage of the different design possibilities. Although still not an exhaustive list, Sussmann and Robertson [1986, p. 462] proposed a classification, shown as Table 3.1 below, of eleven different possible validation designs incorporating the proposals of both Tiffen [1946] and Guion and Cranny [1982].

Table 3.1: Eleven validation research designs variations as a function of timing of test and performance measurement and nature of selection decision.

DESIGN NUMBER	BEFORE EMPLOYMENT	SELECTION DECISION	AT ENTRY	AFTER SHORT TENURE	AFTER EXTENDED TENURE
01	X	R	-	-	P
02	-	R	X	-	P
03	X	E	-	-	P
04	-	E	X	-	P
05	X	X	-	-	P
06	-	R	-	X,P	-
07	-	R	-	-	X,P
08	-	E	-	X,P	-
09	-	E	-	-	X,P
10	-	E	-	-	X,P[CS]
11	-	E	-	-	A,P

Note: X=experimental tests, E=existing tests, R=random selection, P=performance measure, A=archival data and CS=cross-sectional present employees.

Designs 1 to 5 represent predictive designs due to the nontrivial time interval between the collection of the predictor data [X] and the job performance data [P]. Designs 6 to 11, in contrast, represent concurrent designs in view of the fact that the predictor and job performance measures are obtained at [approximately] the same time. The necessity of correcting the obtained validity coefficient for

restriction of range [assuming the initial applicant sample to be representative of the relevant applicant population and assuming no systematic attrition] will depend on [Campbell, 1991; Schepers, 1996; Sussmann & Robertson, 1986]:

- the way in which the selection decision is made [i.e. the way in which the applicants are determined for whom criterion data should be available]; and
- the selection ratio [i.e. the proportion of the initial applicant group for whom criterion data should become available].

If the selection decision is made randomly [designs 1, 2, 6 & 7] no restriction of range corrections would be necessary, provided the initial applicant sample is representative of the relevant applicant population [Huck, 1992; Sussmann & Robertson, 1986;]. Should the selection decision, however, be based on existing selection tests, the necessity of correcting the obtained validity coefficient for restriction of range [still assuming the initial applicant sample to be representative of the relevant applicant population] will depend on the correlation of E with X and P respectively. In the unlikely event of E showing no significant correlation with either X [possible but unlikely] or Y [unlikely because it would imply an invalid current selection procedure], corrections for restriction of range would not be necessary. Should E, however, show a significant correlation with either X or Y, and to the extent that the selection ratio decreases [i.e. smaller proportions of the applicant group are allowed to enter], corrections for restriction of range would be necessary. Similarly, should selection occur on X, and to the extent that the selection ratio decreases, corrections for restriction of range would be necessary. Judged purely in terms of restriction of range as a threat to external validity, there consequently seems to be, contrary to earlier positions [Barrett, Philips & Alexander, 1981; Guion & Cranny, 1982] no justification for any preference for a predictive over a concurrent validation design [Sussmann & Robertson, 1986, p. 464]:

In summary, when restriction of the sample is considered a threat to external validity and to accurate estimation of population validity coefficients, the manner in which the selection decision is made - randomly, existing tests, experimental tests - can, and logically will, make a difference. Predictive and concurrent designs do not differ in respect of this threat because each can be affected by the same factors to the same degree.

This should however not be taken to mean that concurrent validation designs logically constitute an adequate substitute for predictive designs [Guion & Cranny, 1982]. Should the conditions and individuals characterising the simulation of the selection procedure differ in one or more respects from those that would apply at the time of actual application and, should the validation results be sensitive to the features that differ across the areas of analysis and application, the transportability of the research evidence becomes questionable. Table 3.2 compares the area of eventual application of the selection procedure with predictive and concurrent validation designs on a number of features which affect the representativeness of the validation design and thus its ability to successfully substitute for the actual selection procedure.



Table 3.2: Differences in the representativeness of predictive and concurrent validation designs in substituting for the actual selection procedure

DISTINCTIVE FEATURE	PREDICTIVE DESIGN	CONCURRENT DESIGN	ACTUAL SELECTION
X - Y TIME LAG	N	T	N
TEST MOTIVATION	H	L	H
TRAINING EFFECT ON X	L	H	L
EXPERIENCE EFFECT ON X	L	H	L

Note: H=high, L=low, T=trivial time-lag, N=non-trivial time-lag

Differences between a concurrent validation design and the actual area of application on one or more of the features listed in Table 3.2 need not necessarily bias the validation study results obtained through a concurrent design. It does, however, present the opportunity to hypothesise that such differences would distort/bias any one or more of those aspects that the validation study was designed to deliver [i.e. validity coefficient, decision function/regression equation, predictor cut-off/strategy matrix entries, absolute and incremental utility assessments or fairness assessments], thus casting doubt on the legitimacy of the claim to have provided research-based justification for the transportation of those results to the actual area of application. A healthy degree of scientific skepticism/caution is thus called for, should practical constraints effectively rule out the implementation of a proper predictive validation design [Guion & Cranny, 1982].

A further consideration in the argument for correcting the validity coefficient for the effect of restriction of range and criterion unreliability is the reduced power of significance tests applied to the uncorrected correlation coefficient [Sackett & Wade, 1983; Schmidt, Hunter & Urry, 1976; Raju, Edwards & LoVerde, 1985]. Larger samples are thus required to keep power at a specified level [e.g. 0.80] than would have been the case had restriction of range and/or criterion unreliability not attenuated the observed correlation. Corrections to the validity coefficient would, however, not necessarily alleviate the problem since the variance of the sampling distribution of the correlation coefficient is also affected by the corrections [Bobko, 1983; Gross & Kagen, 1983]. Nonetheless, the question whether to correct or not, will not only depend on the effect thereof on the accuracy of the estimate but also on the effect of such corrections on the power of hypothesis tests applied to the corrected correlation coefficient.

The feasibility of implementing the appropriate restriction of range correction formula, should such a correction be a prerequisite to obtain an unbiased estimate of the relevant parametric validity coefficient, firstly depends on whether measurements on the predictor [X], the criterion [Y] or a third variable [Z, which could be an existing selection test or any other measure] are available for both the entire, unrestricted applicant group and the selected or restricted applicant group. Since selection, in the sense with which Table 3.1 uses the term, always occurs on a variable other than the criterion, the probability of Y measurements being available for both the selected and unselected groups for any of

the eleven designs shown in Table 3.1, is zero. Third variable [Z] scores being available depends on whether selection occurred on a third variable and whether the nature of the third variable is known. Predictor [X] scores would be available if:

- selection is directly based on X; or
- the predictor is administered before selection at random or based on a third variable occurs.

The feasibility [or probably more appropriately termed, expediency] of implementing the appropriate restriction of range correction formula, however, is also dependent on the extent to which the linearity and homoscedasticity assumptions underlying the Pearson-Lawley correction formula [Campbell, 1991; Huck, 1992; Olsen & Becker, 1983; Schepers, 1996] can be assessed and the extent to which they are satisfied by the data set.

The assumption of no attrition is clearly unrealistic. In all eleven validation designs attrition would, to the extent that it occurred, necessitate [additional] restriction of range corrections [assuming random attrition to be extremely unlikely]. The appropriate correction should be feasible, as long as it was known whether attrition occurred directly on job performance or on a third variable and Y or Z measurements for both the restricted and unrestricted groups were available. Criterion measurements for both the restricted and unrestricted groups would be logically impossible for any one of the eleven designs in Table 3.1. Criterion measures could, however, be in principle available for pre- and post-attrition groups.

Should support for the performance hypothesis be found [i.e. should an array of significant correlations be found], a description of the functional relationship between the criterion and a [linear] combination of these [significant] predictors need to be inferred to serve as a decision rule for actual personnel selection decisions. In addition descriptive evaluations of the effectiveness, utility and fairness of selection decisions based on the derived decision rule are required. The decision rule and the descriptions of the effectiveness, utility and fairness of selection decisions based on the derived decision rule are, however, just like the validity coefficient, all derived via a simulation of the selection procedure under conditions that differ from those prevailing at the eventual use of the selection procedure. The decision rule and the descriptions of the effectiveness, utility and fairness of selection decisions could, thus, just like the validity coefficient, be biased to such an extent that the transportability thereof becomes problematic. Whether this is in fact the case, and if so, under what conditions and to what extent bias is introduced, is not completely clear although partial details has been published [Linn, 1983; Pearson, 1903;]. Furthermore, appropriate correction formula do not seem to be generally available, but even if they were, the question would still remain whether there is any merit in applying any of these statistical correction formula. The research objectives as outlined in chapter 1 are thus thereby indicated.

### 3.2 THE VALIDITY COEFFICIENT CORRECTED FOR THE ATTENUATING EFFECT OF THE UNRELIABILITY OF THE CRITERION MEASUREMENTS

Random measurement error adversely affects the correlation between variables, thus, diminishing the chances of exposing regularities in nature [Huck, 1992; Lord & Novick, 1968; Nunnally, 1978]. This has already been pointed out by Spearman [1910, p. 271]:

... the apparent degree of correspondence between any two series of measurements is largely affected by the size of the 'accidental' errors in the process of measurement. ... As a remedy, a correction formula was proposed, based on the idea that the size of the 'accidental' errors can be measured by the size of the discrepancies between successive measurements of the same things.

Equation 3.1 gives the correlation between true scores alluded to by Spearman [1910], in terms of the correlation between observed scores and the reliability of each measurement.

Let  $X$  and  $Y$  denote the observed scores on the predictor and criterion respectively. Let  $T_X$ ,  $T_Y$ ,  $E_X$  and  $E_Y$  denote the true and error score components of the observed predictor and criterion scores.

In accordance with Equation 2.1, the true and error score components of the observed predictor and criterion scores combine linearly:

$$X = T_X + E_X \text{-----} 3.1.1$$

And;

$$Y = T_Y + E_Y \text{-----} 3.1.2$$

Solving Equations 3.1.1 and 3.1.2 for  $E_X$  and  $E_Y$  respectively and substituting in the Pearson correlation between the predictor and criterion true scores:

$$\begin{aligned} \rho[T_X, T_Y] &= \sigma[T_X, T_Y] / (\sigma[T_X] \sigma[T_Y]) \\ &= \sigma[(X - E_X), (Y - E_Y)] / (\sigma[T_X] \sigma[T_Y]) \\ &= \{\sigma[X, Y] - \sigma[E_X, Y] - \sigma[X, E_Y] + \sigma[E_X, E_Y]\} / (\sigma[T_X] \sigma[T_Y]) \\ &= \sigma[X, Y] / (\sigma[T_X] \sigma[T_Y]) \text{-----} 3.1.3 \end{aligned}$$

Since,  $\sigma[E_X, Y] = \sigma[X, E_Y] = \sigma[E_X, E_Y] = 0$



However, the correlation between the predictor and criterion observed scores is given by:

$$\rho[X,Y] = \sigma[X,Y]/(\sigma[X]\sigma[Y]) \text{-----} 3.1.4$$

Isolating the covariance term in Equation 3.1.4:

$$\sigma[X,Y] = \rho[X,Y]\sigma[X]\sigma[Y] \text{-----} 3.1.5$$

Substituting Equation 3.1.5 in Equation 3.1.3:

$$\begin{aligned} \rho[T_X,T_Y] &= \rho[X,Y]\sigma[X]\sigma[Y]/(\sigma[T_X]\sigma[T_Y]) \\ &= \rho[X,Y]/\{(\sigma[T_X]/\sigma[X])(\sigma[T_Y]/\sigma[Y])\} \text{-----} 3.1.6 \end{aligned}$$

However, according to Equation 2.14, the correlation between two parallel measures indicate the proportion true score variance, and hence:

$$\sigma[T_X]/\sigma[X] = \sqrt{\rho[X_{ip},X_{iq}]} \text{-----} 3.1.7$$

and

$$\sigma[T_Y]/\sigma[Y] = \sqrt{\rho[Y_{ip},Y_{iq}]} \text{-----} 3.1.8$$

Substituting Equations 3.1.7 and 3.1.8 in Equation 3.1.6:

$$\rho[T_X,T_Y] = \rho[X,Y]/(\sqrt{\rho[X_{ip},X_{iq}]}\sqrt{\rho[Y_{ip},Y_{iq}]}) \text{-----} 3.1.9$$

Write

$$\rho[X_{ip},X_{iq}] = \rho_{ttX}$$

and

$$\rho[Y_{ip},Y_{iq}] = \rho_{ttY}$$

Therefore the fully disattenuated validity coefficient can therefore be written as:

$$\rho[T_X,T_Y] = \rho[X,Y]/(\sqrt{\rho_{ttX}}\sqrt{\rho_{ttY}}) \text{-----} 3.1$$

Equation 3.1 is generally known as a correction/estimation formula for the attenuating effect of the unreliability of two measures on the correlation between their observed scores. Thus, if the reliabilities of the two measures are known, then Equation 3.1 could be used to estimate the so called disattenuated correlation [i.e. the correlation that would have been obtained if the true scores on the two measures would have been correlated]. The disattenuated correlation provided by Equation 3.1 is, however, not appropriate for applied selection validation research.

The correlation of interest in a validation study is the operational validity coefficient; the correlation between the fallible predictor and actual job performance and not some fallible measure of performance. Schmidt, Hunter & Urry [1976, p. 474] explain the necessity of correcting the validity coefficient for the unreliability of the criterion only as follows:

Unreliability in the predictor is 'real' in the sense that, in actual selection use, one must use the test as it exists - unreliability and all. Thus we do not correct the validity coefficient for unreliability in the test; it does no good to compute the validity of a perfect reliable test since we do not have such a test available for use. Unreliability in the criterion on the other hand, is not 'real' in the sense that it does not affect the operational value of the selection test. Once the validation study is completed, the criterion measure is not used further. In making selection decisions we use the test to make predictions of actual job performance, not performance as measured on our imperfect criterion measure. Thus validity coefficients are, or should be, corrected for unreliability in the criterion measure.

Through a procedure similar to that used to derive Equation 3.1, Equation 3.2 is obtained which permits the estimation of the correlation between the observed score on one measure [typically the predictor] and the true score on another measure [the criterion] from the correlation between the observed scores on the two measures and the reliability of the second measure.

Let  $X$  and  $Y$  denote the observed scores on the predictor and criterion respectively. Let  $T_X$ ,  $T_Y$  and  $E_X$  and  $E_Y$  denote the true and error score components of the observed predictor and criterion scores.

In accordance with Equation 2.1, the true and error score components of the observed predictor score combine linearly:

$$X = T_X + E_X \text{-----} 3.2.1$$

Solving Equation 3.2.1 for  $E_X$  and substituting in the Pearson correlation between the observed predictor score and criterion true score:

$$\rho[X, T_Y] = \sigma[X, T_Y] / (\sigma[X] \sigma[T_Y])$$

$$\begin{aligned}
&= \sigma[(T_X + E_X), T_Y] / (\sigma[X]\sigma[T_Y]) \\
&= \{\sigma[T_X, T_Y] + \sigma[T_Y, E_X]\} / (\sigma[X]\sigma[T_Y]) \\
&= \sigma[T_X, T_Y] / (\sigma[X]\sigma[T_Y]) \text{-----} 3.2.2
\end{aligned}$$

Since  $\sigma[T_Y, E_X] = 0$

However:

$$\begin{aligned}
\sigma[X, Y] &= \sigma[(T_X - E_X), (T_Y + E_Y)] \\
&= \sigma[T_X, T_Y] + \sigma[E_X, T_Y] + \sigma[E_Y, T_X] + \sigma[T_X, T_Y] \\
&= \sigma[T_X, T_Y] \text{-----} 3.2.3
\end{aligned}$$

Substituting Equation 3.2.3 in Equation 3.2.2:

$$\begin{aligned}
\rho[X, T_Y] &= \sigma[T_X, T_Y] / (\sigma[X]\sigma[T_Y]) \\
&= \sigma[X, Y] / (\sigma[X]\sigma[T_Y]) \text{-----} 3.2.4
\end{aligned}$$

However, the correlation between the predictor and criterion observed scores is given by:

$$\rho[X, Y] = \sigma[X, Y] / (\sigma[X]\sigma[Y]) \text{-----} 3.2.5$$

Isolating the covariance term in Equation 3.2.5:

$$\sigma[X, Y] = \rho[X, Y]\sigma[X]\sigma[Y] \text{-----} 3.2.6$$

Substituting Equation 3.2.6 in Equation 3.2.4:

$$\begin{aligned}
\rho[X, T_Y] &= \rho[X, Y]\sigma[X]\sigma[Y] / (\sigma[X]\sigma[T_Y]) \\
&= \rho[X, Y]\sigma[Y] / \sigma[T_Y] \\
&= \rho[X, Y] / (\sigma[T_Y] / \sigma[Y]) \\
&= \rho[X, Y] / \sqrt{\rho_{Yip}Y_{iq}} \text{-----} 3.2.7
\end{aligned}$$

Since, according to Equation 2.14, the correlation between two parallel measures indicate the proportion true score variance, and hence:

$$\sigma[T_Y] / \sigma[Y] = \sqrt{\rho_{Yip}Y_{iq}}$$

Write

$$\rho_{YipYiq} \equiv \rho_{ttY}$$

The partially disattenuated validity coefficient can therefore be written as:

$$\rho[X, T_Y] = \rho_{XY} / \sqrt{\rho_{ttY}} \text{-----} 3.2$$

The behaviour of the partially disattenuated correlation coefficient as a function of the attenuated correlation and the reliability coefficient is portrayed graphically in two dimensions in Figure 3.1 and in three dimensions in Figure 3.2. The partially disattenuated correlation coefficient was mapped onto a surface defined by  $0.05 \leq \rho[X, Y] \leq 0.95$  and  $0.05 \leq \rho_{ttY} \leq 1.00$  through a SAS program feeding a sample of surface coordinates into Equation 3.2.

Figures 3.1 and 3.2 indicate that the utility of Equation 3.2 increases as  $\rho_{ttY}$  decreases. The observed correlation coefficient  $\rho[X, Y]$  severely underestimates  $\rho[X, T_Y]$  when the reliability of the criterion is low. At the upper end of the  $\rho_{ttY}$ -scale, however, the degree of underestimation becomes sufficiently small to seriously question the utility of Equation 3.2.

### 3.3 THE VALIDITY COEFFICIENT CORRECTED FOR THE ATTENUATING EFFECT OF EXPLICIT OR IMPLICIT RESTRICTION OF RANGE

#### 3.3.1 Univariate Selection

If a nonzero correlation exists between X and Y within the applicant population, then the correlation coefficient calculated for a random sample drawn from that population would, within the limits of sampling error, approximate the parameter. That is  $E[r_{XY}] \cong \rho[X, Y]$ . If, however, the sample would be selected conditional on the scores of the predictor X or the criterion Y or a third variable Z, then the sample estimate will be systematically negatively biased; that is,  $E[r_{XY}] < \rho[X, Y]$  [Campbell, 1991; Huck, 1992; Olsen & Becker, 1983; Schepers, 1996]. In the typical [predictive] validation study some form of the latter case normally prevails. According to Linn, Harnish and Dunbar [1981], Greener and Osburn [1980], Campbell [1991] and Schepers [1996] the negative bias in the population correlation coefficient estimate created by restriction of range can be quite severe for selection ratios and [unrestricted] validity coefficients that are regularly encountered in validation research.

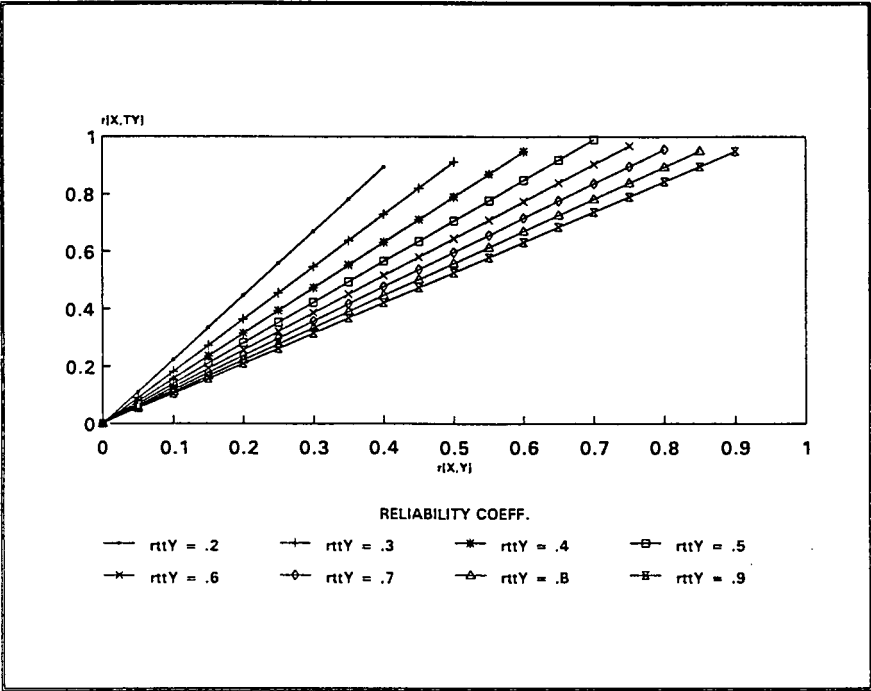


Figure 3.1: The partially disattenuated validity coefficient as a function of  $\rho[X,Y]$  and  $\rho_{ttY}$

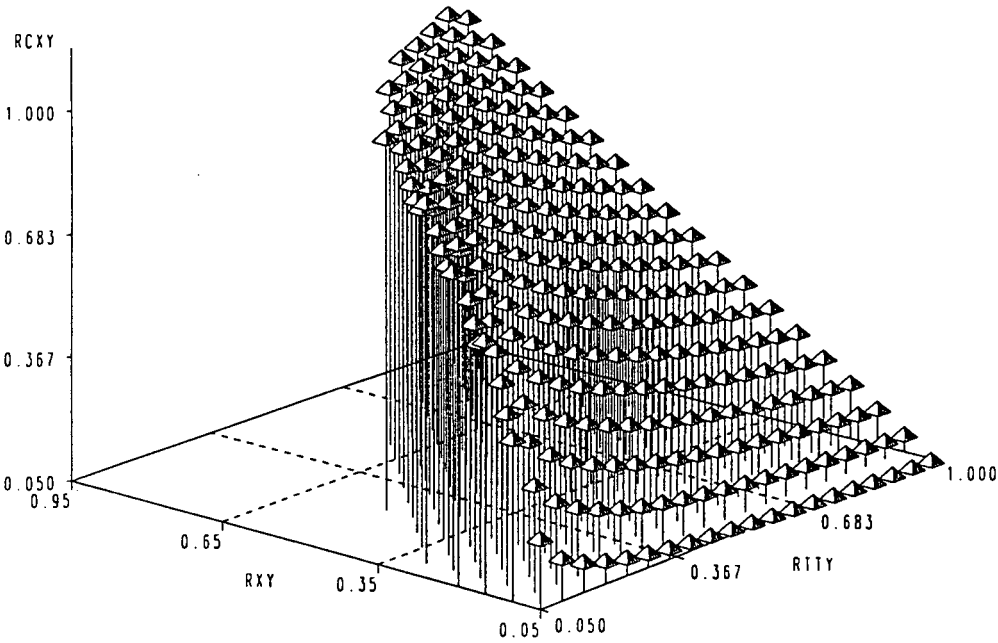


Figure 3.2: The partially disattenuated validity coefficient as a function of  $\rho[X,Y]$  and  $\rho_{ttY}$

All other factors kept constant, the more severe the selection process [i.e. the larger the proportion of the total population with missing values on the incidental selection variable(s)] the stronger the influence of restriction of range [Guion, 1991; Schepers, 1996]. Schepers [1996] reports that very high and very low [unrestricted population] correlations are only slightly influenced by restriction of range.

Correlations in the middle range [i.e. 0,40 - 0,80], however, are severely curtailed by restriction of range [Schepers, 1996]. This would thus imply that the parametric predictive validity of the selection procedure would be substantially misrepresented by the uncorrected sample result. The psychometric problem confronting the validation researcher is to estimate the bivariate correlation for the unrestricted applicant population from the available, biased data.

Pearson [1903] was the first to provide a procedure to correct for the attenuating effect of restriction of range due to explicit or implicit selection. Lawley [1943] relaxed the assumptions required by the Pearson [1903] procedure. Thorndike [1949] is credited with disseminating the Pearson-Lawley procedure among psychometrists [Mendoza, Hart & Powell, 1991].

Considerable confusion and error can be avoided by distinguishing carefully between the different types of selection. Pearson [1903] identified three different conditions under which estimation bias due to selection could occur. From the work of Thorndike [1949; 1982] these three circumstances initially became known as Cases 1, 2 and 3 [Ree, Caretta, Earles & Albert, 1994; Thorndike, 1949]. Thorndike [1982] subsequently also referred to the three conditions under which estimation bias could occur as Cases B, A and C [Olsen & Becker, 1983; Ree, Carretta, Earles & Albert, 1994; Thorndike, 1949; Thorndike, 1982]. These three cases can be distinguished as follows:

- Case 1[Case B]: the correlation to be corrected is between two variables X and Y, selection occurred [directly/explicitly] on the variable X [or Y] through complete truncation on X at  $X_C$  [or on Y at  $Y_C$ ] and both restricted and unrestricted variances are known only for the incidental selection variable Y [or X];
- Case 2 [Case A]: the correlation to be corrected is between two variables X and Y, selection occurred [directly/explicitly] on the variable X [or Y] through complete truncation on X at  $X_C$  [or on Y at  $Y_C$ ] and both restricted and unrestricted variances are known only for the explicit selection variable X [or Y];
- Case 3[i] [Case C[i]]: the correlation to be corrected is between two variables X and Y, selection occurred [directly/explicitly] on the variable Z [a single variable or a composite] through complete truncation on Z at  $Z_C$  [the truncation on Y and X thus is incomplete] and both restricted and unrestricted variances are known only for the explicit selection variable Z; and
- Case 3[ii] [Case C[ii]]: the correlation to be corrected is between two variables X and Y, selection occurred [directly/explicitly] on the variable Z through complete



truncation on Z at  $Z_c$  [the truncation on Y and X thus is incomplete] and both restricted and unrestricted variances are known only for the incidental selection variable X or Y.

### 3.3.1.1 The Effect Of Case 1 [Case B] Restriction Of Range On The Validity Coefficient

In the case of explicit selection on the predictor X [or the criterion Y] and variance known for both the restricted and unrestricted groups on the criterion [or predictor] variable subject to incidental selection [i.e. Case 1 [Case B]], let the to be corrected correlation coefficient calculated for the restricted group be indicated as  $\rho_{[x,y]}$  and the to be estimated correlation coefficient as  $\rho[X,Y]$ . Let  $\sigma^2_{[x]}$  and  $\sigma^2_{[y]}$  represent the calculated [i.e. known] variances for the restricted group and  $\sigma^2[X]$  and  $\sigma^2[Y]$  the variances for the unrestricted group of which only the latter is known. If the selection is explicitly on the predictor X [Cook, Campbell & Peracchio, 1991; Feldt & Brennan, 1989; Ghiselli, Campbell & Zedeck, 1981; Gulliksen, 1950; Thorndike, 1982] and assuming a linear, homoscedastic regression of the criterion Y on the predictor X in the total applicant population:

- the regression of the criterion Y on the predictor X will not be affected [i.e. the mean Y conditional on X is not altered by explicit selection on X although the mean X conditional on Y, and therefore the regression of X on Y, will be altered and vice versa if complete truncation occurred on Y]; and
- the criterion variance conditional on X [i.e. the square of the standard error of estimate,  $\sigma^2[Y|X]$ ] will not be altered.

A correction formula to correct for the selection bias in the correlation coefficient under Case 1 [Case B] conditions [Equation 3.3] can subsequently be derived from these two basic assumptions as indicated below [Gulliksen, 1950; Thorndike, 1982; Ghiselli, Campbell & Zedeck, 1981].

Flowing from the assumption that the regression of Y on X will not be affected by explicit selection on X:

$$\beta[y|x] = \rho_{[x,y]}(\sigma_{[y]}/\sigma_{[x]}) = \beta[Y|X] = \rho[X,Y](\sigma[Y]/\sigma[X]) \text{-----} 3.3.1$$

Flowing from the assumption that the criterion variance conditional on X will not be affected by explicit selection on X:

$$\sigma[y|x] = \sigma[y]\sqrt{(1-\rho^2_{[x,y]})} = \sigma[Y|X] = \sigma[Y]\sqrt{(1-\rho^2[X,Y])} \text{-----} 3.3.2$$

Taking the square root of equation 3.3.2:

$$\sigma^2[y](1-\rho^2[x,y]) = \sigma^2[Y](1-\rho^2[X,Y]) \text{-----} 3.3.3$$

Isolating the term  $(1-\rho^2[X,Y])$  in Equation 3.3.3:

$$\sigma^2[y]/\sigma^2[Y](1-\rho^2[x,y]) = (1-\rho^2[X,Y]) \text{-----} 3.3.4$$

Isolating the squared unrestricted correlation coefficient in Equation 3.3.4:

$$\rho^2[X,Y] = 1 - \{(\sigma^2[y]/\sigma^2[Y])(1-\rho^2[x,y])\} \text{-----} 3.3.5$$

Taking the square root of Equation 3.3.5:

$$\begin{aligned} \rho[X,Y] &= \sqrt{(1 - \{(\sigma^2[y]/\sigma^2[Y])(1-\rho^2[x,y])\})} \\ &= \sqrt{(\sigma^2[Y] - \sigma^2[y] + \sigma^2[y]\rho^2[x,y])/\sigma^2[Y]} \\ &= \sqrt{(1 - K^2(1-\rho^2[x,y]))} \text{-----} 3.3 \end{aligned}$$

Where  $K = \sigma[y]/\sigma[Y]$

The reaction of the correlation corrected for Case 1 [Case B] restriction of range to changes in K and the magnitude of the obtained, restricted correlation is graphically portrayed in Figure 3.3 and Figure 3.4. The correlation coefficient corrected for Case 1 [Case B] selection on X was mapped onto a surface defined by  $0.01 \leq \rho[x,y] \leq 1.00$  and  $0.10 \leq K \leq 1.00$  through a SAS program feeding a sample of surface coordinates into Equation 3.3.

Figures 3.3 and 3.4 indicate that the amount of benefit derived from Equation 3.3 increases as the severity of selection increases [i.e. as K decreases]. Figures 3.3 and 3.4 also indicate that the effect of Case 1 [Case B] restriction of range decreases as  $\rho[X,Y]$  increases.

It should be noted that  $\sigma^2[Y]$  need not be greater than  $\sigma^2[y]$  as was assumed in the preceding derivation. Equation 3.3 would still apply if the lower-case subscripts are interpreted to designate the group for which complete information [i.e.  $\rho[x,y]$ ,  $\sigma[x]$  and  $\sigma[y]$ ] is available [Gulliksen, 1950].

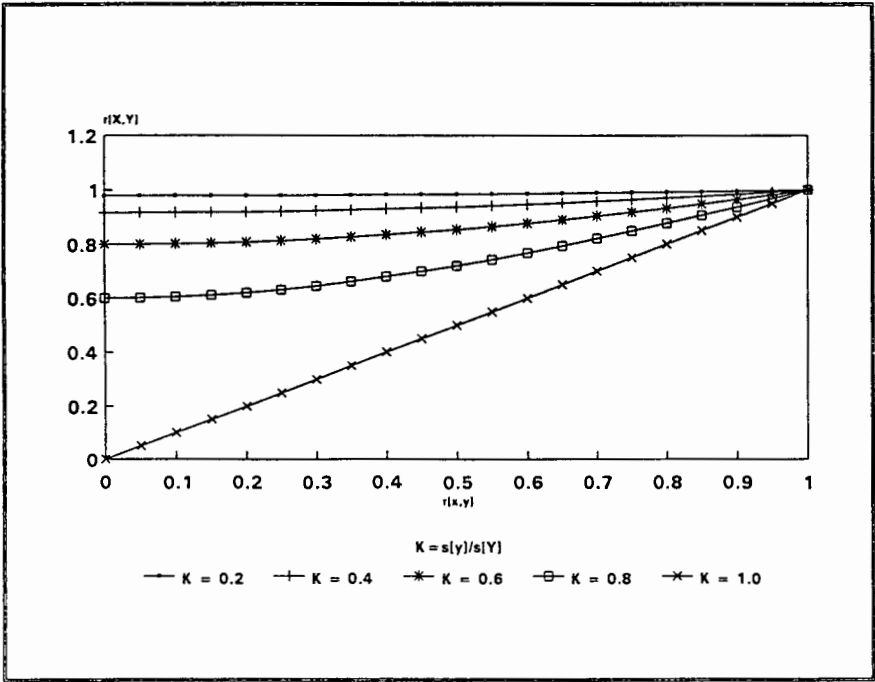


Figure 3.3: The validity coefficient corrected for Case 1 [Case B] restriction of range as a function of  $K$  and  $\rho[x,y]$

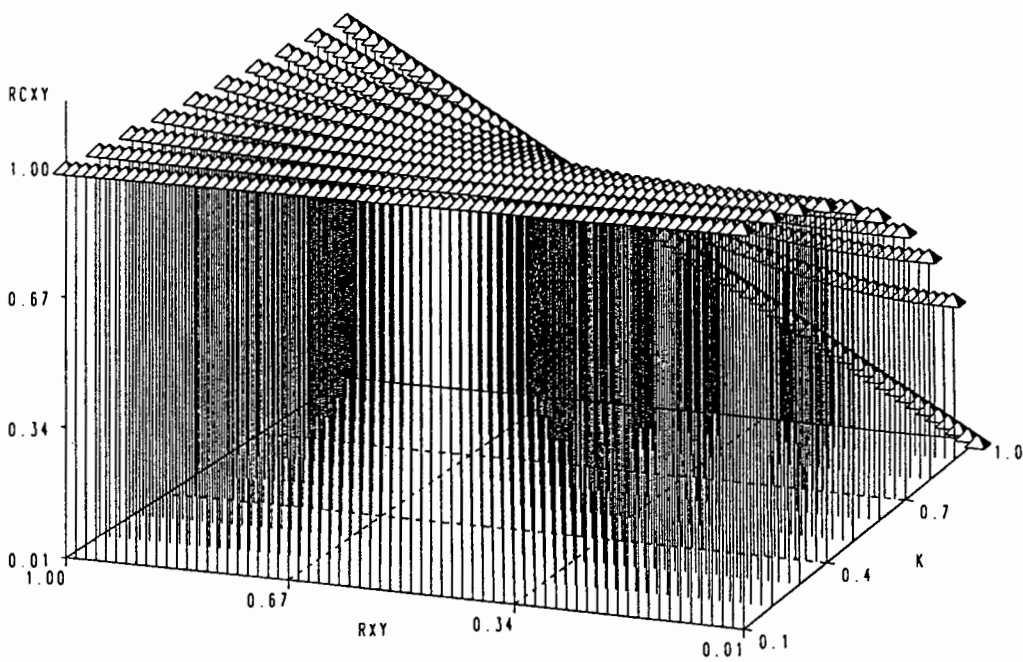


Figure 3.4: The validity coefficient corrected for Case 1 [Case B] restriction of range as a function of  $K$  and  $\rho[x,y]$

### 3.3.1.2 The Effect Of Case 2 [Case A] Restriction Of Range On The Validity Coefficient

In the more common case of explicit selection on the predictor X and variance known for both the restricted and unrestricted groups on the predictor variable subject to direct selection [i.e. Case 2 [Case A]], let the to be corrected correlation coefficient calculated for the restricted group again be indicated as  $\rho[x,y]$  and the to be estimated correlation coefficient still as  $\rho[X,Y]$ . Let  $\sigma^2[x]$  and  $\sigma^2[y]$  represent the calculated [i.e. known] variances for the restricted group and  $\sigma^2[X]$  and  $\sigma^2[Y]$  the variances for the unrestricted group of which only the former is known.

If the selection is explicitly on the predictor X the same two assumptions applicable to Case 1 [Case B] also apply to Case 2 [Case A]. A correction formula to correct for the selection bias in the correlation coefficient under Case 2 [Case A] conditions [Equation 3.4] can subsequently be derived from these two basic assumptions as indicated below [Gulliksen, 1950; Lord & Novick, 1968; Ghiselli, Campbell & Zedeck, 1981; Thorndike, 1982].

Flowing from the assumption that the regression of Y on X will not be affected by explicit selection on X:

$$\beta[y|x] = \rho[x,y](\sigma[y]/\sigma[x]) = \beta[Y|X] = \rho[X,Y](\sigma[Y]/\sigma[X]) \text{-----} 3.4.1$$

Flowing from the assumption that the criterion variance conditional on X will not be affected by explicit selection on X:

$$\sigma[y|x] = \sigma[y]\sqrt{(1-\rho^2[x,y])} = \sigma[Y|X] = \sigma[Y]\sqrt{(1-\rho^2[X,Y])} \text{-----} 3.4.2$$

Isolating the term  $(\sigma[Y]/\sigma[X])$  in Equation 3.4.1:

$$\sigma[Y]/\sigma[X] = (\rho[x,y]\sigma[y]) / (\sigma[x]\rho[X,Y]) \text{-----} 3.4.3$$

Isolating the unrestricted criterion standard deviation in Equation 3.4.3:

$$\sigma[Y] = (\rho[x,y]\sigma[y]\sigma[X]) / (\sigma[x]\rho[X,Y]) \text{-----} 3.4.4$$

Taking the square root of Equation 3.4.2 and substituting the unrestricted criterion standard deviation in Equation 3.4.2 with Equation 3.4.4:

$$\sigma[y](1-\rho^2[xy])^{1/2} = \sigma[Y](1-\rho^2[X,Y])^{1/2}$$

$$= \{(\rho[x,y]\sigma[y]\sigma[X])/(\sigma[x]\rho[X,Y])\}(1-\rho^2[X,Y])^{1/2} \text{-----} 3.4.5$$

Squaring Equation 3.4.5:

$$\sigma^2[y](1-\rho^2[xy]) = \{(\rho^2[x,y]\sigma^2[y]\sigma^2[X])/(\sigma^2[x]\rho^2[X,Y])\}(1-\rho^2[X,Y]) \text{-----} 3.4.6$$

Isolating the term  $(1-\rho^2[xy])$  in Equation 3.4.6:

$$(1-\rho^2[xy]) = \{(\rho^2[x,y]\sigma^2[X])/(\sigma^2[x]\rho^2[X,Y])\}(1-\rho^2[X,Y]) \text{-----} 3.4.7$$

Cross multiplying Equation 3.4.7:

$$(1-\rho^2[xy])(\sigma^2[x]\rho^2[X,Y]) = (\rho^2[x,y]\sigma^2[X])(1-\rho^2[X,Y]) \text{-----} 3.4.8$$

Multiplying Equation 3.4.8 with  $1/(\rho^2[x,y]\sigma^2[X])(1/\rho^2[X,Y])$ :

$$\begin{aligned} \sigma^2[x](1-\rho^2[xy])/(\rho^2[x,y]\sigma^2[X]) &= (1-\rho^2[X,Y])/\rho^2[X,Y] \\ &= (1/\rho^2[X,Y]) - 1 \text{-----} 3.4.9 \end{aligned}$$

Simplifying the left-hand term in Equation 3.4.9:

$$\begin{aligned} (1/\rho^2[X,Y]) - 1 &= \sigma[x](1-\rho^2[xy])/(\rho^2[x,y]\sigma^2[X]) \\ &= (\sigma^2[x] - \sigma^2[x]\rho^2[xy])/(\rho^2[x,y]\sigma^2[X]) \\ &= (\sigma^2[x]/\rho^2[x,y] - \sigma^2[x])/ \sigma^2[X] \\ &= (\sigma^2[x]/\sigma^2[X])(1/\rho^2[x,y] - 1) \text{-----} 3.4.10 \end{aligned}$$

Isolating the term  $(1/\rho^2[X,Y])$  in Equation 3.4.10, multiplying the right-hand side out and creating  $(\sigma^2[X]\rho^2[x,y])$  as a common denominator:

$$\begin{aligned} (1/\rho^2[X,Y]) &= 1 + (\sigma^2[x]/\sigma^2[X])(1/\rho^2[x,y] - 1) \\ &= 1 + \sigma^2[x]/(\sigma^2[X]\rho^2[x,y]) - (\sigma^2[x]/\sigma^2[X]) \\ &= (\sigma^2[X]\rho^2[x,y] + \sigma^2[x] - \sigma^2[x]\rho^2[x,y])/(\sigma^2[X]\rho^2[x,y]) \text{-----} 3.4.11 \end{aligned}$$

Isolating the square of the unrestricted correlation coefficient in Equation 3.4.11 through an inversion of Equation 3.4.11:

$$\begin{aligned} \rho^2[X,Y] &= 1/\{(\sigma^2[X]\rho^2[x,y] + \sigma^2[x] - \sigma^2[x]\rho^2[x,y])/(\sigma^2[X]\rho^2[x,y])\} \\ &= (\sigma^2[X]\rho^2[x,y])/(\sigma^2[X]\rho^2[x,y] + \sigma^2[x] - \sigma^2[x]\rho^2[x,y]) \text{-----} 3.4.12 \end{aligned}$$

Taking the square root of Equation 3.4.12 and multiplying denominator and :numerator with  $(1/\sigma[x])$

$$\begin{aligned}\rho[X,Y] &= (\sigma[X]\rho[x,y]) / (\sigma^2[X]\rho^2[x,y] + \sigma^2[x] - \sigma^2[x]\rho^2[x,y])^{1/2} \\ &= (\sigma[X]/\sigma[x])\rho[x,y] / \{(\sigma^2[X]/\sigma^2[x])\rho^2[x,y] + (\sigma^2[x]/\sigma^2[x]) - (\sigma^2[x]/\sigma^2[x])\rho^2[x,y]\}^{1/2} \\ &= (\sigma[X]/\sigma[x])\rho[x,y] / \{(\sigma^2[X]/\sigma^2[x])\rho^2[x,y] + 1 - \rho[x,y]\}^{1/2} \\ &= K\rho[x,y] / (K^2\rho^2[x,y] + 1 - \rho^2[x,y])^{1/2} \text{-----} 3.4\end{aligned}$$

Where:

$$K = \sigma[X]/\sigma[x]$$

The reaction of the obtained, restricted correlation under Case 2 [Case A] selection to changes in  $K$  and  $\rho[x,y]$  is illustrated graphically in Figure 3.5 and Figure 3.6. Figures 3.5 and 3.6 confirms the position of Schepers [1996, p. 22] that "both very high and very low correlations are only slightly influenced by restriction of range, but the middle range of correlations [0,4 to 0,8] are strongly influenced." Figures 3.5 and 3.6 also indicate that the effect of Case 2 [Case A] restriction of range increases as  $K$  increases.

Case 2 [Case A] has a reasonable practical significance because it is encountered whenever an estimate of the validity of a selection device is desired that has actually been used to select the group for which criterion scores become available.

### 3.3.1.3 The Effect Of Case C Restriction Of Range On The Validity Coefficient

Under the third possible condition of explicit selection on a third variable  $Z$  and variance known for both the restricted and unrestricted groups on the third variable subject to direct selection [i.e. Case 3[i] [Case C[i]]], let the to be corrected correlation coefficient calculated for the restricted group again be indicated as  $\rho[x,y]$  and the to be estimated correlation coefficient still as  $\rho[X,Y]$ .

Let  $\sigma^2[x]$ ,  $\sigma^2[y]$  and  $\sigma^2[z]$  represent the calculated [i.e. known] variances and  $\rho[x,z]$  and  $\rho[y,z]$  the calculated correlation coefficients for the restricted group and  $\sigma^2[X]$ ,  $\sigma^2[Y]$  and  $\sigma^2[Z]$  the variances for the unrestricted group of which only the latter is known. If the selection is explicitly on the third variable  $Z$  the following five assumptions need to be made in addition to the assumption of linear, homoscedastic regression in the total applicant population.

- the regression of the criterion  $Y$  on the third variable  $Z$  will not be affected [i.e. the mean  $Y$  conditional on  $Z$  is not altered by explicit selection on  $Z$ ;



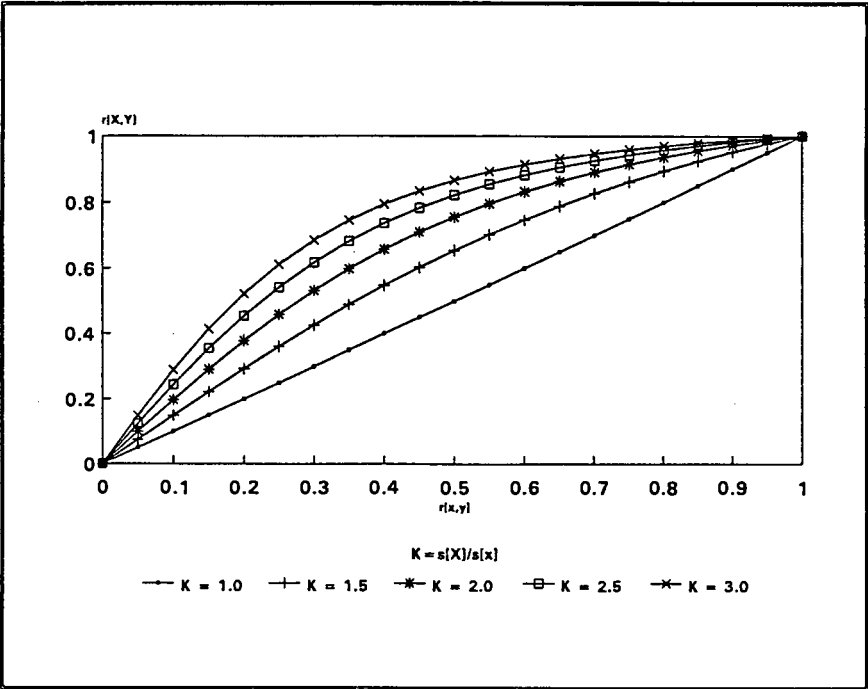


Figure 3.5: The validity coefficient corrected for Case 2 [Case A] restriction of range as a function of  $K$  and  $\rho[x,y]$

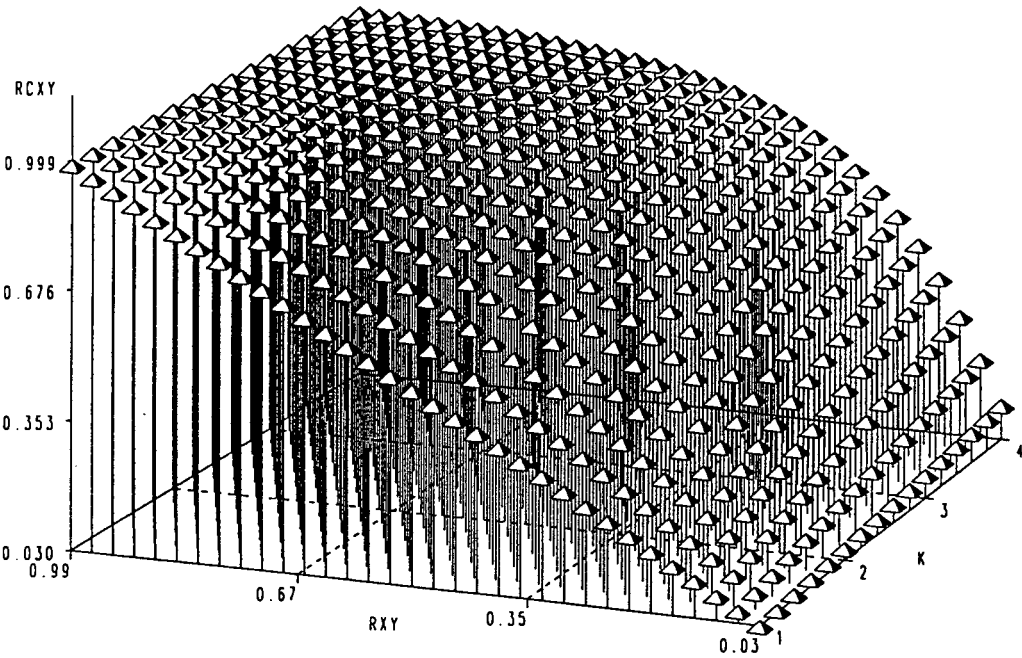


Figure 3.6: The validity coefficient corrected for Case 2 [Case A] restriction of range as a function of  $K$  and  $\rho[x,y]$

- the regression of the predictor X on the third variable Z will not be affected [i.e. the mean X conditional on Z is not altered by explicit selection on Z;
- the criterion variance conditional on Z [i.e. the square of the standard error of estimate,  $\sigma^2[Y|Z]$  will not be altered;
- the X variance conditional on Z [i.e. the square of the standard error of estimate,  $\sigma^2[X|Z]$  will not be altered; and
- the correlation between the predictor X and the criterion Y for a constant Z [i.e. the partial correlation  $\rho[X,Y.Z]$ ] will not be altered by explicit selection on Z.

A correction formula to correct for the selection bias in the correlation coefficient under Case 3[i] [Case C[i]] conditions [Equation 3.5] can subsequently be derived from these basic assumptions as indicated below [Gulliksen, 1950; Lord & Novick, 1968; Ghiselli, Campbell & Zedeck, 1981; Thorndike, 1982].

Flowing from the the assumption that the regression of Y on Z is not affected by explicit selection on Z:

$$\rho[y,z](\sigma[y]/\sigma[z]) = \rho[Y,Z](\sigma[Y]/\sigma[Z]) \text{-----} 3.5.1$$

Flowing from the assumption that the regression of X on Z is not affected by explicit selection on Z:

$$\rho[x,z](\sigma[x]/\sigma[z]) = \sigma[X,Z](\sigma[X]/\sigma[Z]) \text{-----} 3.5.2$$

Flowing from the assumption that the criterion variance conditional on Z will not be affected by explicit selection on Z:

$$\sigma[y]\sqrt{(1 - \rho^2[y,z])} = \sigma[Y]\sqrt{(1 - \rho^2[Y,Z])} \text{-----} 3.5.3$$

Flowing from the assumption that the predictor variance conditional on Z will not be affected by explicit selection on Z:

$$\sigma[x]\sqrt{(1 - \rho^2[x,z])} = \sigma[X]\sqrt{(1 - \rho^2[X,Z])} \text{-----} 3.5.4$$

Flowing from the assumption that the correlation between the predictor and the criterion with Z held constant will not be affected by explicit selection on Z:

$$\begin{aligned} \rho[x,y.z] &= (\rho[x,y]-\rho[x,z]\rho[y,z])/\sqrt{\{(1-\rho^2[x,z])(1-\rho^2[y,z])\}} \\ &= \rho[X,Y.Z] \\ &= (\rho[X,Y]-\rho[X,Z]\rho[Y,Z])/\sqrt{\{(1-\rho^2[X,Z])(1-\rho^2[Y,Z])\}} \text{-----} 3.5.5 \end{aligned}$$

Isolating the term  $\rho[Y,Z]$  in Equation 3.5.1:

$$\rho[Y,Z] = \rho[y,z](\sigma[y]\sigma[Z]) / (\sigma[z]\sigma[Y]) \text{-----} 3.5.6$$

Substituting Equation 3.5.6 in Equation 3.5.3:

$$\begin{aligned} \sigma[y]\sqrt{(1 - \rho^2[y,z])} &= \sigma[Y]\sqrt{(1 - \rho^2[Y,Z])} \\ &= \sigma[Y]\sqrt{(1 - \rho^2[y,z](\sigma^2[y]\sigma^2[Z]) / (\sigma^2[z]\sigma^2[Y]))} \text{-----} 3.5.7 \end{aligned}$$

Squaring Equation 3.5.7:

$$\begin{aligned} \sigma^2[y](1 - \rho^2[y,z]) &= \sigma^2[Y](1 - \rho^2[y,z](\sigma^2[y]\sigma^2[Z]) / (\sigma^2[z]\sigma^2[Y])) \\ &= \sigma^2[Y] - \rho^2[y,z](\sigma^2[y]\sigma^2[Z]) / \sigma^2[z] \text{-----} 3.5.8 \end{aligned}$$

Isolating the unrestricted criterion variance term in Equation 3.5.8:

$$\begin{aligned} \sigma^2[Y] &= \sigma^2[y](1 - \rho^2[y,z]) + \rho^2[y,z](\sigma^2[y]\sigma^2[Z]) / \sigma^2[z] \\ &= \sigma^2[y]\{1 - \rho^2[y,z] + \rho^2[y,z](\sigma^2[Z] / \sigma^2[z])\} \text{-----} 3.5.9 \end{aligned}$$

Substituting 3.5.9 in Equation 3.5.6:

$$\begin{aligned} \rho[Y,Z] &= \rho[y,z](\sigma[y]\sigma[Z]) / (\sigma[z]\sigma[Y]) \\ &= \rho[y,z](\sigma[y]\sigma[Z]) / (\sigma[z]\sigma[y]\sqrt{\{1 - \rho^2[y,z] + \rho^2[y,z](\sigma^2[Z] / \sigma^2[z])\}}) \\ &= \rho[y,z](\sigma[Z] / \sigma[z]) / \sqrt{\{1 - \rho^2[y,z] + \rho^2[y,z](\sigma^2[Z] / \sigma^2[z])\}} \text{-----} 3.5.10 \end{aligned}$$

Through a similar derivation it can be shown that:

$$\sigma^2[X] = \sigma^2[x]\{1 - \rho^2[x,z] + \rho^2[x,z](\sigma^2[Z] / \sigma^2[z])\} \text{-----} 3.5.11$$

and;

$$\rho[X,Z] = \rho[x,z](\sigma[Z] / \sigma[z]) / \sqrt{\{1 - \rho^2[x,z] + \rho^2[x,z](\sigma^2[Z] / \sigma^2[z])\}} \text{-----} 3.5.12$$

From Equations 3.5.3 and 3.5.4 it follows that:

$$\sqrt{(1 - \rho^2[Y,Z])} = (\sigma[y] / \sigma[Y])\sqrt{(1 - \rho^2[y,z])} \text{-----} 3.5.13$$

and:

$$\sqrt{(1 - \rho^2[X,Z])} = (\sigma[x]/\sigma[X])\sqrt{(1 - \rho^2[x,z])} \text{-----} 3.5.14$$

Substituting Equations 3.5.13 and 3.5.14 in Equation 3.5.5:

$$\begin{aligned} & (\rho[x,y]-\rho[x,z]\rho[y,z])/\sqrt{(1-\rho^2[x,z])(1-\rho^2[y,z])} \\ &= (\rho[X,Y]-\rho[X,Z]\rho[Y,Z])/(\sigma[x]/\sigma[X])\sqrt{(1-\rho^2[x,z])}(\sigma[y]/\sigma[Y])\sqrt{(1-\rho^2[y,z])} \\ &= (\rho[X,Y]-\rho[X,Z]\rho[Y,Z])(\sigma[Y]\sigma[X])/\{\sqrt{(1-\rho^2[x,z])}\sqrt{(1-\rho^2[y,z])}\sigma[y]\sigma[x]\} \text{-----} 3.5.15 \end{aligned}$$

Cross multiplying Equation 3.5.15:

$$\begin{aligned} & (\rho[x,y]-\rho[x,z]\rho[y,z])\{\sqrt{(1 - \rho^2[x,z])}\sqrt{(1 - \rho^2[y,z])}(\sigma[y]\sigma[x])\} \\ &= (\rho[X,Y]-\rho[X,Z]\rho[Y,Z])(\sigma[Y]\sigma[X])\sqrt{\{(1-\rho^2[x,z])(1-\rho^2[y,z])\}} \text{-----} 3.5.16 \end{aligned}$$

Multiplying Equation 3.5.16 with  $1/\{\sqrt{(1-\rho^2[x,z])}(1-\rho^2[y,z])\}$ :

$$\begin{aligned} & (\rho[X,Y]-\rho[X,Z]\rho[Y,Z])(\sigma[Y]\sigma[X]) \\ &= \{(\rho[x,y]-\rho[x,z]\rho[y,z])\{\sqrt{(1-\rho^2[x,z])}\sqrt{(1-\rho^2[y,z])}\sigma[y]\sigma[x]\}\}/\sqrt{\{(1-\rho^2[x,z])(1-\rho^2[y,z])\}} \\ &= (\rho[x,y]-\rho[x,z]\rho[y,z])(\sigma[y]\sigma[x]) \text{-----} 3.5.17 \end{aligned}$$

Multiplying Equation 3.5.17 with  $1/(\sigma[Y]\sigma[X])$ :

$$(\rho[X,Y]-\rho[X,Z]\rho[Y,Z]) = (\rho[x,y]-\rho[x,z]\rho[y,z])(\sigma[y]\sigma[x])/(\sigma[Y]\sigma[X]) \text{-----} 3.5.18$$

Isolating the unrestricted predictor-criterion correlation in Equation 3.5.18:

$$\rho[X,Y] = (\rho[x,y]-\rho[x,z]\rho[y,z])(\sigma[y]\sigma[x])/(\sigma[Y]\sigma[X]) + \rho[X,Z]\rho[Y,Z] \text{-----} 3.5.19$$

Substituting Equations 3.5.10 and 3.5.12 in Equation 3.5.19:

$$\begin{aligned} \rho[X,Y] &= (\rho[x,y]-\rho[x,z]\rho[y,z])(\sigma[y]\sigma[x])/(\sigma[Y]\sigma[X]) + \rho[X,Z]\rho[Y,Z] \\ &= (\rho[x,y]-\rho[x,z]\rho[y,z])(\sigma[y]\sigma[x])/(\sigma[Y]\sigma[X]) + \\ &\quad \{\rho[x,z](\sigma[Z]/\sigma[z])/\sqrt{(1-\rho^2[x,z])} + \\ &\quad \rho^2[x,z]\sigma^2[Z]/\sigma^2[z]\}\{\rho[y,z](\sigma[Z]/\sigma[z])/\sqrt{(1-\rho^2[y,z])} + \rho^2[y,z]\sigma^2[Z]/\sigma^2[z]\} \\ &= (\rho[x,y]-\rho[x,z]\rho[y,z])(\sigma[y]\sigma[x])/(\sigma[Y]\sigma[X]) + \\ &\quad \{\rho[x,z]\rho[y,z](\sigma^2[Z]/\sigma^2[z])\}/\{\sqrt{(1-\rho^2[x,z])} + \rho^2[x,z]\sigma^2[Z]/\sigma^2[z]\}\sqrt{(1-\rho^2[y,z])} + \\ &\quad \rho^2[y,z]\sigma^2[Z]/\sigma^2[z]\} \text{-----} 3.5.20 \end{aligned}$$

Substituting Equation 3.5.9 and 3.5.11 in Equation 3.5.20:

$$\begin{aligned}
 \rho[X,Y] &= (\rho[x,y]-\rho[x,z]\rho[y,z])(\sigma[y]\sigma[x]) / \{(\sigma[y]\{1-\rho^2[y,z]+ \\
 &\quad \rho^2[y,z]\sigma^2[Z]/\sigma^2[z]\})\{\sigma[x](1-\rho^2[x,z]+ \\
 &\quad \rho^2[x,z]\sigma^2[Z]/\sigma^2[z])^{1/2}\}\} + \\
 &\quad \{\rho[x,z]\rho[y,z](\sigma^2[Z]/\sigma^2[z])\} / \{\sqrt{(1-\rho^2[x,z]+\rho^2[x,z]\sigma^2[Z]/\sigma^2[z])}\sqrt{(1-\rho^2[y,z]+ \\
 &\quad \rho^2[y,z]\sigma^2[Z]/\sigma^2[z])}\} \\
 &= \{(\rho[x,y]-\rho[x,z]\rho[y,z]) + \\
 &\quad \{\rho[x,z]\rho[y,z](\sigma^2[Z]/\sigma^2[z])\}\} / \{\sqrt{(1-\rho^2[x,z]+\rho^2[x,z]\sigma^2[Z]/\sigma^2[z])}\sqrt{(1-\rho^2[y,z]+ \\
 &\quad \rho^2[y,z]\sigma^2[Z]/\sigma^2[z])}\} \text{-----} 3.5.21
 \end{aligned}$$

Write:

$$\rho[X,Y] = (\rho[x,z]\rho[y,z]K^2) / \{\sqrt{(1-\rho^2[x,z]+\rho^2[x,z]K^2)}\sqrt{(1-\rho^2[y,z]+\rho^2[y,z]K^2)}\} \text{-----} 3.5$$

Where:

$$K = (\sigma[Z]/\sigma[z])$$

The behaviour of the correlation coefficient corrected for Case 3[i] [Case C[i]] restriction of range is shown in Figures 3.7 - 3.12. The correlation coefficient corrected for Case 3[i] [Case C[i]] selection on Z was mapped onto a surface defined by K,  $\rho[x,z]$ ,  $\rho[y,z]$  and  $\rho[x,y]$  through a SAS program feeding a sample of surface coordinates into Equation 3.5. Figures 3.7 - 3.12 indicate that the utility of Equation 3.5 increases as K increases [i.e. the selection ratio decreases] and the correlation of the selection variable Z with the criterion and the predictor increases.

Finally, under the condition of explicit selection on a third variable Z and variance known for both the restricted and unrestricted groups on either the predictor or criterion variables subject to incidental selection [i.e. Case 3[ii] [Case C[ii]]], let the to be corrected correlation coefficient calculated for the restricted group again be indicated as  $\rho[x,y]$  and the to be estimated correlation coefficient still as  $\rho[X,Y]$ . Let  $\sigma^2[x]$ ,  $\sigma^2[y]$  and  $\sigma^2[z]$  represent the calculated [i.e. known] variances and  $\rho[x,z]$  and  $\rho[y,z]$  the known correlation coefficients for the restricted group. Let  $\sigma^2[X]$ ,  $\sigma^2[Y]$  and  $\sigma^2[Z]$  represent the variances for the unrestricted group of which either the first or the second is known. If the selection is explicitly on the third variable Z the same five assumptions need to be made in addition to the assumption of linear, homoscedastic regression in the total applicant population as for Case C [i]. A correction formula to correct for the selection bias in the correlation coefficient under Case 3[ii] [Case C[ii]] conditions [Equation 3.6] can subsequently be derived from these basic assumptions as indicated below [Gulliksen, 1950].

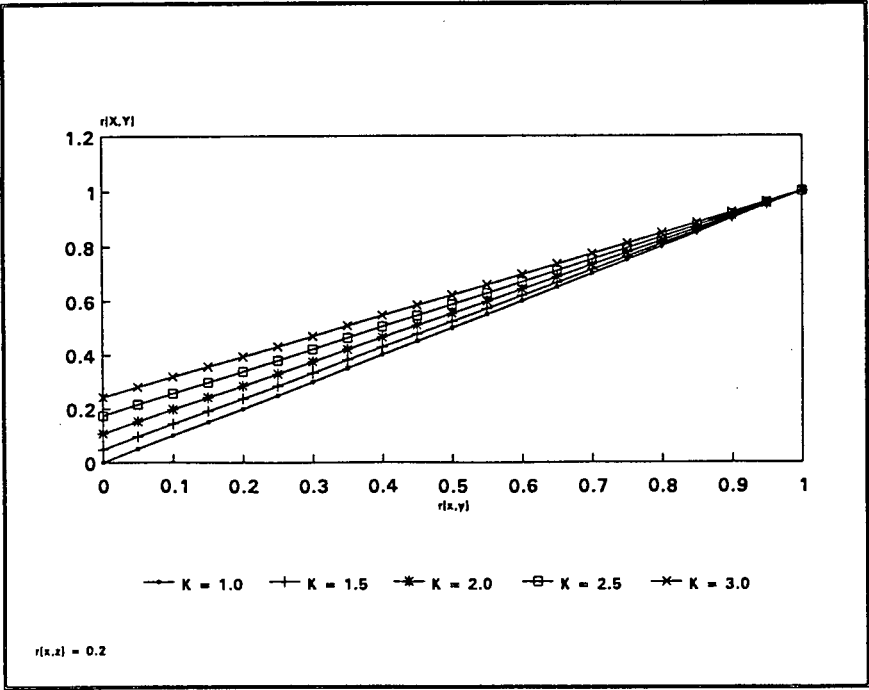


Figure 3.7: The validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of  $K$ ,  $\rho[y,z]$  and  $\rho[x,y]$  for  $\rho[x,z]$  fixed at 0.2

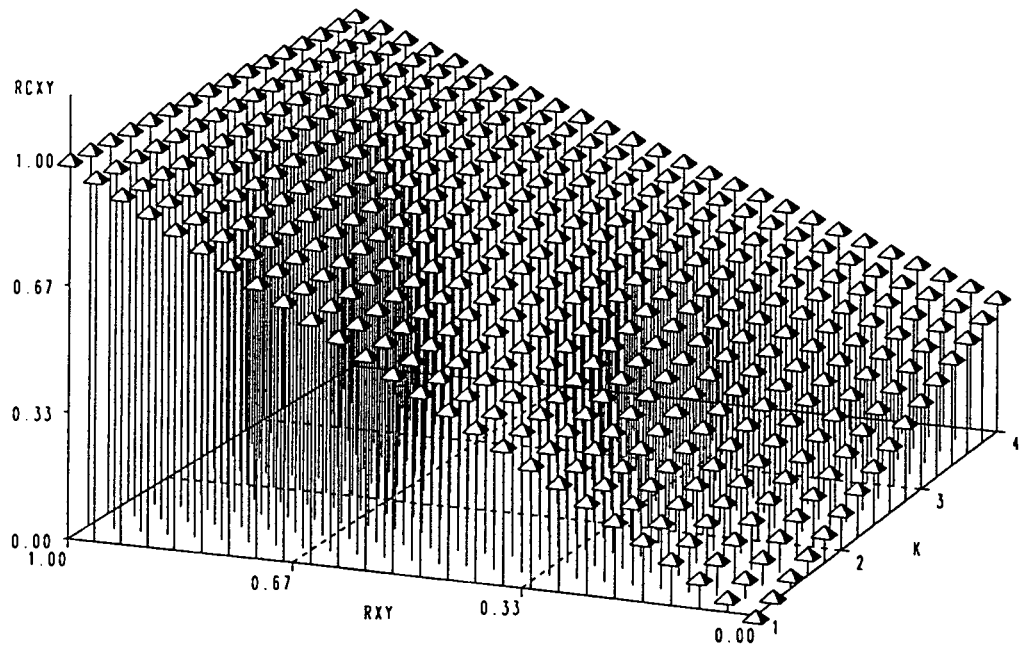


Figure 3.8: The validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of  $K$ ,  $\rho[y,z]$  and  $\rho[x,y]$  for  $\rho[x,z]$  fixed at 0.2



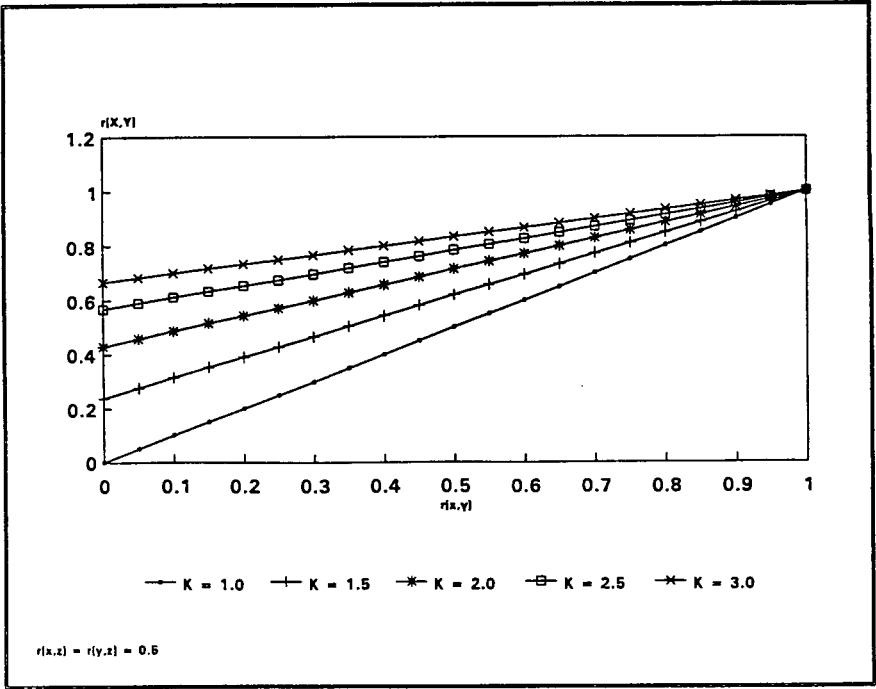


Figure 3.9: The validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of  $K$ ,  $\rho[y,z]$  and  $\rho[x,y]$  for  $\rho[x,z]$  fixed at 0.5

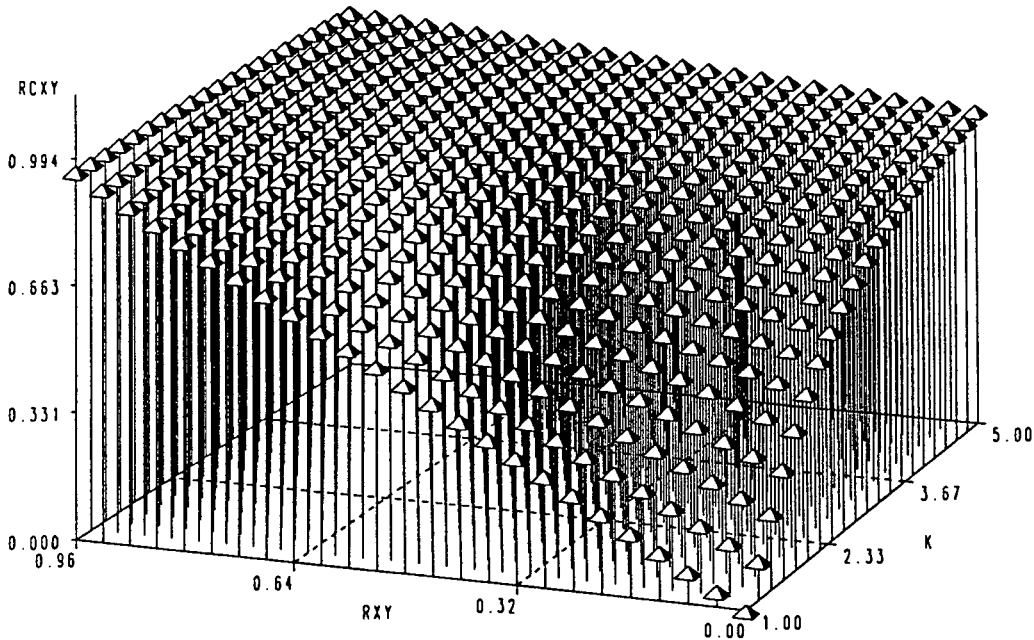


Figure 3.10: The validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of  $K$ ,  $\rho[y,z]$  and  $\rho[x,y]$  for  $\rho[x,z]$  fixed at 0.5

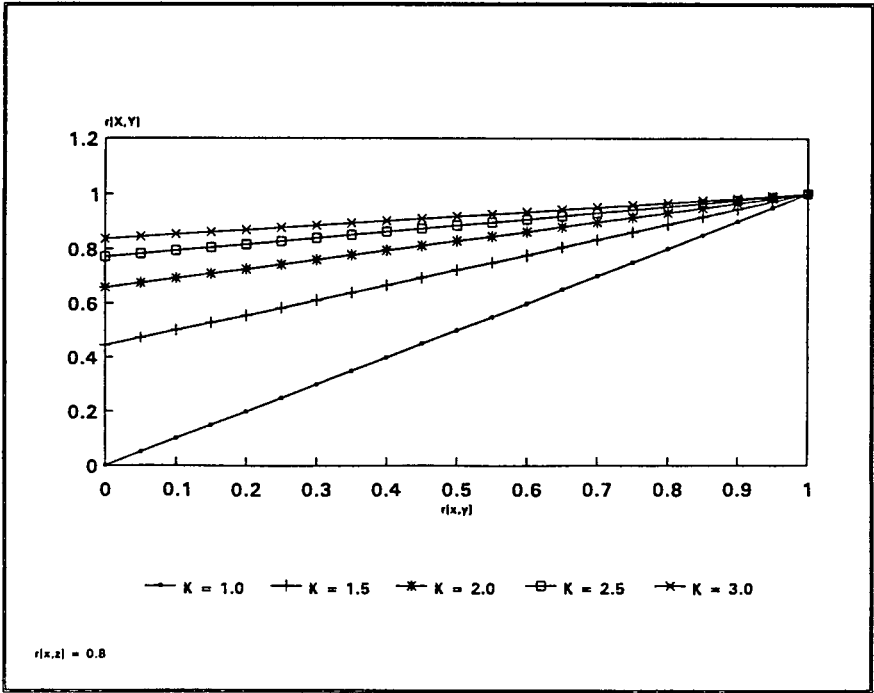


Figure 3.11: The validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of  $K$ ,  $\rho[y,z]$  and  $\rho[x,y]$  for  $\rho[x,z]$  fixed at 0.8

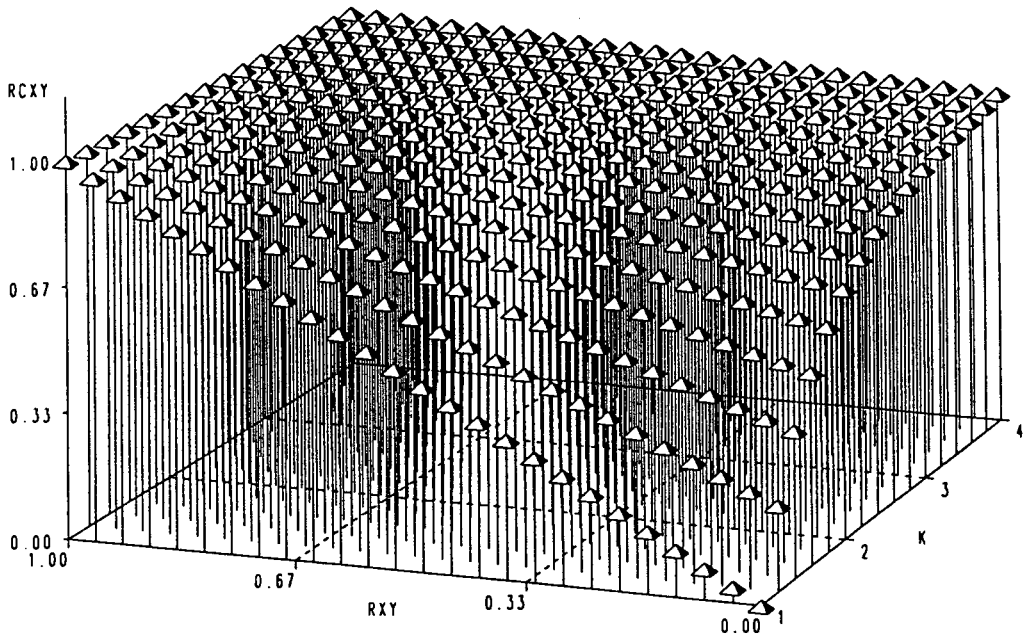


Figure 3.12: The validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of  $K$ ,  $\rho[y,z]$  and  $\rho[x,y]$  for  $\rho[x,z]$  fixed at 0.8

Flowing from the the assumption that the regression of Y on Z is not affected by explicit selection on Z:

$$\rho_{[y,z]}(\sigma_{[y]}/\sigma_{[z]}) = \rho[Y,Z](\sigma[Y]/\sigma[Z]) \text{-----} 3.6.1$$

Flowing from the assumption that the regression of X on Z is not affected by explicit selection on Z:

$$\rho_{[x,z]}(\sigma_{[x]}/\sigma_{[z]}) = \rho[X,Z](\sigma[X]/\sigma[Z]) \text{-----} 3.6.2$$

Flowing from the assumption that the criterion variance conditional on Z will not be affected by explicit selection on Z:

$$\sigma_{[y]}\sqrt{(1 - \rho^2_{[y,z]})} = \sigma[Y]\sqrt{(1 - \rho^2[Y,Z])} \text{-----} 3.6.3$$

Flowing from the assumption that the predictor variance conditional on Z will not be affected by explicit selection on Z:

$$\sigma_{[x]}\sqrt{(1 - \rho^2_{[x,z]})} = \sigma[X]\sqrt{(1 - \rho^2[X,Z])} \text{-----} 3.6.4$$

Flowing from the assumption that the correlation between the predictor and the criterion with Z held constant will not be affected by explicit selection on Z:

$$\begin{aligned} \rho_{[x,y,z]} &= (\rho_{[x,y]}\rho_{[x,z]}\rho_{[y,z]})/\sqrt{\{(1-\rho^2_{[x,z]})(1-\rho^2_{[y,z]})\}} \\ &= \rho[X,Y,Z] \\ &= (\rho[X,Y]\rho[X,Z]\rho[Y,Z])/\sqrt{\{(1-\rho^2[X,Z])(1-\rho^2[Y,Z])\}} \text{-----} 3.6.5 \end{aligned}$$

Squaring Equation 3.6.4:

$$\sigma^2_{[x]}(1 - \rho^2_{[x,z]}) = \sigma^2[X](1 - \rho^2[X,Z]) \text{-----} 3.6.6$$

Isolating the term  $\rho^2[X,Z]$  in Equation 3.6.6:

$$\rho^2[X,Z] = 1 - (\sigma^2_{[x]}/\sigma^2[X])(1 - \rho^2_{[x,z]}) \text{-----} 3.6.7$$

Cross multiplying Equation 3.6.2:

$$\rho_{[x,z]}\sigma_{[x]}\sigma_{[Z]} = \rho[X,Z]\sigma[X]\sigma_{[z]} \text{-----} 3.6.8$$

Isolating the term  $\sigma[Z]/\sigma[z]$  in Equation 3.6.8 and substituting Equation 3.6.7:

$$\begin{aligned}
 \sigma[Z]/\sigma[z] &= (\rho[X,Z]\sigma[X]) / (\rho[x,z]\sigma[x]) \\
 &= \{ \{ 1 - (\sigma^2[x]/\sigma^2[X])(1 - \rho^2[x,z]) \}^{1/2} \sigma[X] \} / (\rho[x,z]\sigma[x]) \\
 &= \{ 1 - (\sigma[x]/\sigma[X]) \{ (1 - \rho^2[x,z]) \}^{1/2} \sigma[X] \} / (\rho[x,z]\sigma[x]) \\
 &= \{ \sigma[X] - \sigma[x] \{ (1 - \rho^2[x,z]) \}^{1/2} \} / (\rho[x,z]\sigma[x]) \\
 &= \{ \sigma^2[X] - \sigma^2[x](1 - \rho^2[x,z]) \}^{1/2} / (\rho[x,z]\sigma[x]) \text{-----} 3.6.9
 \end{aligned}$$

Isolating the unrestricted third variable variance in Equation 3.6.9:

$$\sigma[Z] = \sigma[z] \{ \{ \sigma^2[X] - \sigma^2[x](1 - \rho^2[x,z]) \}^{1/2} / (\rho[x,z]\sigma[x]) \} \text{-----} 3.6.10$$

Isolating the term  $\rho[Y,Z]$  in Equation 3.6.1:

$$\rho[Y,Z] = \rho[y,z] \{ (\sigma[y]\sigma[Z]) / (\sigma[z]\sigma[Y]) \} \text{-----} 3.6.11$$

Squaring Equation 3.6.3 and substituting Equation Equation 3.6.11 in Equation 3.6.3:

$$\begin{aligned}
 \sigma^2[y](1 - \rho^2[y,z]) &= \sigma^2[Y](1 - \rho^2[Y,Z]) \\
 &= \sigma^2[Y] \{ 1 - \rho^2[y,z] \{ (\sigma^2[y]\sigma^2[Z]) / (\sigma^2[z]\sigma^2[Y]) \} \} \\
 &= \sigma^2[Y] - \rho^2[y,z] \{ (\sigma^2[y]\sigma^2[Z]) / \sigma^2[z] \} \text{-----} 3.6.12
 \end{aligned}$$

Isolating the unrestricted criterion variance term in Equation 3.6.12 and substituting Equation 3.6.10 in Equation 3.6.12:

$$\begin{aligned}
 \sigma^2[Y] &= \sigma^2[y](1 - \rho^2[y,z]) + \rho^2[y,z] \{ (\sigma^2[y]\sigma^2[Z]) / \sigma^2[z] \} \\
 &= \sigma^2[y] \{ (1 - \rho^2[y,z]) + \rho^2[y,z] (\sigma^2[Z] / \sigma^2[z]) \} \\
 &= \sigma^2[y] \{ (1 - \rho^2[y,z]) + \\
 &\quad \rho^2[y,z] \{ \{ \sigma^2[z] \{ \sigma^2[X] - \\
 &\quad \sigma^2[x](1 - \rho^2[x,z]) / (\rho^2[x,z]\sigma^2[x]) \} \} / \sigma^2[z] \} \} \text{-----} 3.6.13
 \end{aligned}$$

By canceling out the term  $\sigma^2[z]$  in Equation 3.6.13 and multiplying out the inside brackets:

$$\begin{aligned}
 \sigma^2[Y] &= \sigma^2[y] \{ \{ 1 - \rho^2[y,z] + \rho^2[y,z] \{ \{ \sigma^2[X] - \sigma^2[x](1 - \rho^2[x,z]) \} / (\rho^2[x,z]\sigma^2[x]) \} \} \} \\
 &= \sigma^2[y] \{ \{ \rho^2[x,z]\sigma^2[x] - \rho^2[y,z]\rho^2[x,z]\sigma^2[x] + \\
 &\quad \rho^2[y,z]\sigma^2[X] - \sigma^2[x]\rho^2[y,z] + \\
 &\quad \rho^2[x,z]\rho^2[y,z]\sigma^2[x] \} / \rho^2[x,z]\sigma^2[x] \} \\
 &= \sigma^2[y] \{ \{ \rho^2[x,z]\sigma^2[x] + \rho^2[y,z]\sigma^2[X] - \sigma^2[x]\rho^2[y,z] \} / \rho^2[x,z]\sigma^2[x] \} \text{-----} 3.6.14
 \end{aligned}$$

However, Equations 3.6.3 and 3.6.4, combined permit the following expression:

$$\{\sigma[Y]\sqrt{(1-\rho^2[Y,Z])}\}/\{\sigma[x]\sqrt{(1-\rho^2[x,z])}\} = \{\sigma[y]\sqrt{(1-\rho^2[y,z])}\}/\{\sigma[X]\sqrt{(1-\rho^2[X,Z])}\} \text{-----} 3.6.15$$

Cross multiplying Equation 3.6.15:

$$\sigma[Y]\sigma[X]\sqrt{(1-\rho^2[Y,Z])}\sqrt{(1-\rho^2[X,Z])} = \sigma[y]\sigma[x]\sqrt{(1-\rho^2[y,z])}\sqrt{(1-\rho^2[x,z])} \text{-----} 3.6.16$$

Multiplying Equation 3.6.15 with  $1/(\sigma[Y]\sigma[X])$ :

$$\sqrt{(1-\rho^2[Y,Z])}\sqrt{(1-\rho^2[X,Z])} = \{\sigma[y]\sigma[x]\sqrt{(1-\rho^2[y,z])}\sqrt{(1-\rho^2[x,z])}\}/(\sigma[Y]\sigma[X]) \text{-----} 3.6.17$$

Substituting Equation 3.6.17 in Equation 3.6.5:

$$\begin{aligned} \rho[x,y,z] &= (\rho[x,y]-\rho[x,z]\rho[y,z])/\{\sqrt{(1-\rho^2[x,z])}\sqrt{(1-\rho^2[y,z])}\} \\ &= \rho[X,Y,Z] \\ &= (\rho[X,Y]-\rho[X,Z]\rho[Y,Z])/\{\sqrt{(1-\rho^2[X,Z])}\sqrt{(1-\rho^2[Y,Z])}\} \\ &= (\rho[X,Y]-\rho[X,Z]\rho[Y,Z])/\{\{\sigma[y]\sigma[x]\sqrt{(1-\rho^2[y,z])}\sqrt{(1-\rho^2[x,z])}\}/(\sigma[Y]\sigma[X])\} \\ &= \{(\rho[X,Y]-\rho[X,Z]\rho[Y,Z])(\sigma[Y]\sigma[X])\}/\{\sigma[y]\sigma[x]\sqrt{(1-\rho^2[y,z])}\sqrt{(1-\rho^2[x,z])}\} \text{-----} 3.6.18 \end{aligned}$$

Cross multiplying Equation 3.6.18:

$$\begin{aligned} (\rho[x,y]-\rho[x,z]\rho[y,z])\{\sigma[y]\sigma[x]\sqrt{(1-\rho^2[y,z])}\sqrt{(1-\rho^2[x,z])}\} \\ = \sqrt{(1-\rho^2[x,z])}\sqrt{(1-\rho^2[y,z])}(\rho[X,Y]-\rho[X,Z]\rho[Y,Z])(\sigma[Y]\sigma[X]) \text{-----} 3.6.19 \end{aligned}$$

Multiplying Equation 3.6.19 with  $1/\{\sqrt{(1-\rho^2[x,z])}\sqrt{(1-\rho^2[y,z])}\}$ :

$$\begin{aligned} (\rho[X,Y]-\rho[X,Z]\rho[Y,Z])(\sigma[Y]\sigma[X]) \\ = \{(\rho[x,y]-\rho[x,z]\rho[y,z])\{\sigma[y]\sigma[x]\sqrt{(1-\rho^2[y,z])}\sqrt{(1-\rho^2[x,z])}\}\}/\{\sqrt{(1-\rho^2[x,z])}\sqrt{(1-\rho^2[y,z])}\} \\ = (\rho[x,y]-\rho[x,z]\rho[y,z])(\sigma[y]\sigma[x]) \text{-----} 3.6.20 \end{aligned}$$

Multiplying Equation 3.6.20 with  $1/(\sigma[Y]\sigma[X])$ :

$$\rho[X,Y]-\rho[X,Z]\rho[Y,Z] = (\rho[x,y]-\rho[x,z]\rho[y,z])(\sigma[y]\sigma[x])/(\sigma[Y]\sigma[X]) \text{-----} 3.6.21$$

Isolating the unrestricted predictor-criterion correlation in Equation 3.6.21:

$$\rho[X,Y] = (\rho[x,y]-\rho[x,z]\rho[y,z])(\sigma[y]\sigma[x])/(\sigma[Y]\sigma[X]) + \rho[X,Z]\rho[Y,Z] \text{-----} 3.6.22$$

From Equations 3.6.1 and 3.6.2 it follows that:

$$\{\rho[y,z](\sigma[y]/\sigma[z])\} / \{\rho[X,Z](\sigma[X]/\sigma[Z])\} = \{\rho[Y,Z](\sigma[Y]/\sigma[Z])\} / \{\rho[x,z](\sigma[x]/\sigma[z])\} \text{-----} 3.6.23$$

Cross multiplying Equation 3.6.23:

$$\begin{aligned} \{\rho[X,Z](\sigma[X]/\sigma[Z])\} \{\rho[Y,Z](\sigma[Y]/\sigma[Z])\} &= \{\rho[y,z](\sigma[y]/\sigma[z])\} \{\rho[x,z](\sigma[x]/\sigma[z])\} \\ &= \rho[y,z]\rho[x,z]\{(\sigma[y]\sigma[x])/(\sigma[z]\sigma[z])\} \text{-----} 3.6.24 \end{aligned}$$

Therefore:

$$\rho[X,Z]\rho[Y,Z]\{(\sigma[X]\sigma[Y])/(\sigma^2[Z])\} = \rho[y,z]\rho[x,z]\{(\sigma[y]\sigma[x])/(\sigma^2[z])\} \text{-----} 3.6.25$$

Multiplying Equation 3.6.25 with  $\sigma^2[Z]/(\sigma[X]\sigma[Y])$

$$\rho[X,Z]\rho[Y,Z] = \rho[y,z]\rho[x,z]\{(\sigma[y]\sigma[x]\sigma^2[Z])/(\sigma^2[z]\sigma[X]\sigma[Y])\} \text{-----} 3.6.26$$

Substituting Equation 3.6.26 in Equation 3.6.22:

$$\begin{aligned} \rho[X,Y] &= \{(\rho[x,y]-\rho[x,z]\rho[y,z])(\sigma[y]\sigma[x])/(\sigma[Y]\sigma[X])\} + \\ &\quad \rho[y,z]\rho[x,z]\{(\sigma[y]\sigma[x]\sigma^2[Z])/(\sigma^2[z]\sigma[X]\sigma[Y])\} \\ &= \{(\sigma[y]\sigma[x])/(\sigma[Y]\sigma[X])\}(\rho[x,y]-\rho[x,z]\rho[y,z]) + \\ &\quad \rho[y,z]\rho[x,z](\sigma^2[Z]/\sigma^2[z]) \text{-----} 3.6.27 \end{aligned}$$

Substituting for  $\sigma[Y]$  and  $\sigma[Z]$  in Equation 3.6.27 through Equations 3.6.14 and 3.6.10:

$$\begin{aligned} \rho[X,Y] &= (\sigma[y]\sigma[x]) / \{ \{ \sigma[y] \{ (\rho^2[x,z]\sigma^2[x] + \\ &\quad \rho^2[y,z]\sigma^2[X] - \sigma^2[x]\rho^2[y,z])^{1/2} / \rho[x,z]\sigma[x] \} \} \sigma[X] \} \{ \rho[x,y] - \rho[x,z]\rho[y,z] \} + \\ &\quad \rho[y,z]\rho[x,z] \{ \{ \sigma[z] \{ \{ \sigma^2[X] - \sigma^2[x](1 - \rho^2[x,z]) \}^{1/2} / \rho[x,z]\sigma[x] \} \} / \sigma^2[z] \} \\ &= \{ \rho[x,z]\sigma^2[x]\rho[x,y] + \rho[y,z](\sigma^2[X] - \sigma^2[x]) \} / \{ \sigma[X](\rho^2[x,z]\sigma^2[x] + \\ &\quad \rho^2[y,z](\sigma^2[X] - \sigma^2[x]))^{1/2} \} \text{-----} 3.6 \end{aligned}$$

Case C, in which restriction is indirect, imposed by explicit selection on a third [single or composite] variable, seems to be the most common and most important case in personnel validation research.

Restriction of range through systematic selection reduces the magnitude of the correlation estimate, sometimes to such an extent that the sign of the correlation calculated for the selected group differs in sign from that applicable to the unrestricted group [Linn, 1983; Ree, Carretta, Earles & Albert, 1994].



However, not all of the procedures suggested by Pearson [1903] allow the sign of the corrected correlation to change. Case C [i & ii] selection and corrections are the only univariate cases where the sign of the corrected correlation can differ from the sign of the restricted correlation [Linn, 1983; Ree, Carretta, Earles & Albert, 1994]. According to Ree, Carretta, Earles and Albert [1994] the probability of sign change increases as the correlation between the variable subjected to explicit selection [Z] and the two variables subjected to incidental selection [X and Y], increases. A second factor that influences change in the sign of the correlation coefficient, according to Ree, Carretta, Earles and Albert [1994], is the ratio of the unrestricted to restricted variances which in turn is a function of the selection ratio and, for the variables subjected to incidental selection, the magnitude of the correlation between them and the selection variable. Thus the more severe the selection effect the greater the potential for sign change. The fact that only Case C allows the sign of the corrected correlation to differ from the sign of the restricted correlation implies that the application of an inappropriate univariate correction formula can lead to severely biased estimates of the unrestricted correlation coefficient.

### 3.3.2 Multivariate Selection

Restriction of range in the variables being correlated can result with more than three variables involved. Case C actually represents a special multivariate case of just a single explicit selector variable and only two incidental selector variables [Held & Foley, 1994]. Under a generalisation of Case C a multivariate correction is appropriate. Lawley [1943] developed a solution to the multivariate problem of correcting for restriction of range resulting from selection on an array of variables [Held & Foley, 1994; Ree, Carretta, Earles & Albert, 1994].

The equations needed to correct for the effect of multivariate selection become practically unmanageably complex unless matrix algebra is used [Gulliksen, 1950; Lord & Novick, 1968;]. Multivariate selection will, however, not be formally considered.

### 3.3.3 Accuracy Of Corrections To The Validity Coefficient For The Attenuating Effect Of Explicit Or Implicit Restriction Of Range

Accuracy of estimation is usually defined in terms of the criterion of bias. The central question, thus is the extent to which the corrected correlation coefficient deviates from the actual unrestricted coefficient [i.e.  $E[\hat{\rho}[X,Y] - \rho[X,Y]]$ ] [Gross & Kagen, 1983]. A number of basic issues have relevance

for the question on the accuracy of the Pearson-Lawley corrections to the validity coefficient for the attenuating effect of restriction of range when viewed from the perspective of estimation bias.

A first issue concerns the extent to which the assumptions of linearity and homoscedasticity are violated. Greener and Osburn [1979] showed that departures from linearity tend to deflate the corrected correlation, whilst lack of homoscedasticity tends to have the opposite effect. When the regression of the criterion  $Y$  on the predictor  $X$  is not linear,  $\beta[Y|X] \neq \beta[y|x]$ , and when the regression of the criterion  $Y$  on the predictor  $X$  is not linear and/or the conditional variance of  $Y$  given  $X$  is a function of  $X$ ,  $\sigma[Y|X] \neq \sigma[y|x]$ . Greener and Osburn [1979; 1980] showed that the corrected estimates are, however, still less biased than the uncorrected estimates over a wide range of violations. Greener and Osburn [1979; 1980] also report the corrected estimates to be fairly robust with regards to violations of the assumption of homoscedasticity but more sensitive to departures of linearity. Gross [1982], however, shows that the univariate correction formula [Gulliksen, 1950; Lord & Novick, 1968; Thorndike, 1982] can be valid even when the regression of  $Y$  on  $X$  is non-linear and the conditional  $Y$ -variances are not constant for all values of  $X$ . According to Gross [1982] a simple sufficient condition for the validity of the Pearson correction formula can be stated in terms of the quantity  $Q$  shown as Equation 3.7 below. Gross [1982, p. 798] argues that  $Q = 1$  represents a sufficient condition for the univariate Pearson correction formula to return an unbiased estimate of the unrestricted correlation coefficient:

$$Q = \{(\sigma^2[Y|X]/\sigma^2[y|x])^{1/2}\}/(\beta[Y|X]/\beta[y|x]) \text{-----} 3.7$$

If it is assumed that  $Q = 1$ , then:

$$\begin{aligned} & \{\rho[X,Y](\sigma[Y]/\sigma[X])\}/\{\rho[x,y](\sigma[y]/\sigma[x])\} \\ &= \{\sigma[Y](1-\rho^2[X,Y])^{1/2}\}/\{\sigma[y](1-\rho^2[x,y])^{1/2}\} \text{-----} 3.7.1 \end{aligned}$$

Multiplying Equation 3.7.1 with  $(\sigma[y]/\sigma[Y])$ :

$$\begin{aligned} & \{\rho[X,Y](\sigma[Y]/\sigma[X])\}/\{\rho[x,y](\sigma[y]/\sigma[x])\}(\sigma[y]/\sigma[Y]) \\ &= \{\sigma[Y](1-\rho^2[X,Y])^{1/2}\}/\{\sigma[y](1-\rho^2[x,y])^{1/2}\}(\sigma[y]/\sigma[Y]) \text{-----} 3.7.2 \end{aligned}$$

Simplifying Equation 3.7.2:

$$\{(\rho[X,Y]/\rho[x,y])(\sigma[x]/\sigma[X])\} = (1 - \rho^2[X,Y])^{1/2}/(1 - \rho^2[x,y])^{1/2} \text{-----} 3.7.3$$

Squaring Equation 3.7.3:

$$\{(\rho^2[X,Y]/\rho^2[x,y])(\sigma^2[x]/\sigma^2[X])\} = (1 - \rho^2[X,Y])/(1 - \rho^2[x,y]) \text{-----} 3.7.4$$

Subtracting  $\{(\rho^2[X,Y]/\rho^2[x,y])(\sigma^2[x]/\sigma^2[X])\}$  on both sides of Equation 3.7.4:

$$(1 - \rho^2[X,Y])/(1 - \rho^2[x,y]) - \{(\rho^2[X,Y]/\rho^2[x,y])(\sigma^2[x]/\sigma^2[X])\} = 0 \text{-----} 3.7.5$$

Multiplying Equation 3.7.5 with  $(1 - \rho^2[x,y])$ :

$$(1 - \rho^2[X,Y]) - (1 - \rho^2[x,y])\{(\rho^2[X,Y]/\rho^2[x,y])(\sigma^2[x]/\sigma^2[X])\} = 0 \text{-----} 3.7.6$$

Multiplying Equation 3.7.6 with  $(-1)$ :

$$(-1)\{(1 - \rho^2[X,Y]) - (1 - \rho^2[x,y])\{(\rho^2[X,Y]/\rho^2[x,y])(\sigma^2[x]/\sigma^2[X])\}\} = 0 \text{-----} 3.7.7$$

Therefore:

$$\rho^2[X,Y] + (1 - \rho^2[x,y])\{(\rho^2[X,Y]/\rho^2[x,y])(\sigma^2[x]/\sigma^2[X])\} = 1 \text{-----} 3.7.8$$

By factoring out the term  $\rho^2[X,Y]$  in Equation 3.7.8:

$$\rho^2[X,Y]\{1 + (1 - \rho^2[x,y])(\sigma^2[x]/(\rho^2[x,y]\sigma^2[X]))\} = 1 \text{-----} 3.7.9$$

Placing the term  $\{1 + (1 - \rho^2[x,y])(\sigma^2[x]/(\rho^2[x,y]\sigma^2[X]))\}$  in Equation 3.7.9 on a common denominator:

$$\rho^2[X,Y]\{(\rho^2[x,y]\sigma^2[X] + (1 - \rho^2[x,y])\sigma^2[x]) / (\rho^2[x,y]\sigma^2[X])\} = 1 \text{-----} 3.7.10$$

Therefore:

$$\rho^2[X,Y]\{(\rho^2[x,y] + (1 - \rho^2[x,y])(\sigma^2[x]/\sigma^2[X])) / \rho^2[x,y]\} = 1 \text{-----} 3.7.11$$

Cross multiplying Equation 3.7.11 and isolating the term  $\rho^2[X,Y]$ :

$$\rho^2[X,Y] = \rho^2[x,y] / \{\rho^2[x,y] + (1 - \rho^2[x,y])(\sigma^2[x]/\sigma^2[X])\} \text{-----} 3.7.12$$

Taking the square root of Equation 3.7.12:

$$\begin{aligned} \rho[X,Y] &= \rho[x,y] / \{\rho^2[x,y] + (1 - \rho^2[x,y])(\sigma^2[x]/\sigma^2[X])\}^{1/2} \\ &\equiv \text{Equation 3.4 if } Q = 1 \end{aligned}$$

The traditional linearity and homoscedasticity assumptions thus represent only a special case where  $Q = 1$  and consequently where the correction formula value equals  $\rho[X,Y]$  [Gross, 1982; Held & Foley, 1992]. However, the condition  $Q = 1$  may still hold even if the regression is non-linear and heteroscedastic, although not necessarily for all selection [i.e. restriction] procedures [Gross, 1982].

Conversely, assuming  $\beta[Y|X]$  and  $\beta[y|x]$  are positive, the condition  $Q > 1$  would imply that the correction formula will overestimate the unrestricted correlation coefficient, whereas  $Q < 1$  would imply an underestimation of the unrestricted correlation coefficient [i.e. the validity coefficient pertaining to the unselected applicant population] [Gross, 1982; Held & Foley, 1992; Schepers, 1996]. The condition  $Q > 1$  [i.e. the corrected value shows a positive bias] will result only when the regression of  $Y$  on  $X$  is non-linear and heteroscedastic in the unrestricted group, and selection occurs such that the restricted group has a steeper slope and smaller  $\sigma[y|x]$  than the unrestricted group. Campbell [1991] maintains that the conditions under which this would happen are unlikely to be encountered in practice. Holmes [1990] contends that overestimation will occur only when the ratio of the standard error of prediction at the mean is at least three times as great as at the upper extreme and the selection ratio is 0,40 or less. Again, such a diamond-shaped bivariate scatterplot seems highly improbable [Campbell, 1991]. Schepers [1996] views Gross'  $Q$ -coefficient as an important contribution of great practical value and recommends its routine application whenever corrections for restriction of range are made. Obtaining all the relevant data to calculate  $Q$  might, however, often present an insurmountable obstacle to the implementation of this recommendation.

A second issue concerns the extent to which all the variables on which selection was based are in fact acknowledged as explicit selection variables in the correction formula. In essence the question is thus whether the chosen formula provides an adequate model of the selection process. Quite often the reality of the selection process is not as neat and simple to fit into one of the categories of the Thorndike [1949; 1982] taxonomy. The most important problem seems to be that selection in reality most probably would have been based partly [i.e. in addition to an objectively measured single or composite selection variable] or totally on a number of unspecified variables combined in an unspecifiable way [i.e. clinically combining clinical judgements]. Consequently, and contrary to Case 1 [Case B] and Case 2 [Case A] assumptions, incomplete truncation on the selection variable occur [Olsen & Becker, 1983]. The real problem, however, lies in the fact that measurements of the true explicit selection variables are not accessible and consequently, the Pearson-Lawley correction formula presented thus far, offer no solution to the problem. Thorndike [1949, p. 176] expresses the dilemma experienced by the preceding, traditional correction formula as follows:

[The aforementioned situation] represents an insuperable obstacle to any analytical treatment. When selection is based, as it often is, on clinical judgement which combines in an unspecified and inconstant fashion various types of data about the applicant, and when this judgement is not expressed in any type of



quantitative score, one is at a loss as how to estimate the extent to which the validity coefficient for any test procedure has been affected by that screening.

According to Campbell [1991] selection on variables not accounted for by the chosen correction formula will bias the correction in ways that are at this point in time not altogether predictable. At the same time, however, Campbell [1991] maintains that selection on variables not accounted for in the correction formula never produces positive bias in the corrected coefficient. Levine [1972] and McGuire [1986] contend that the corrected estimate might actually be lower than the uncorrected correlation coefficient if the unknown selection variable is a very strong suppressor variable [i.e. highly correlated with X but uncorrelated with Y] or the distributions are extremely leptokurtic. The probability of these radical conditions actually being encountered are, however, extremely remote [Campbell, 1991; Levine, 1972; McGuire, 1986]. Treating an incidental selection variable as if it were an explicit selection variable typically results in a too conservative correction [Lord & Novick, 1968]. Lord and Novick [1968, p. 148] furthermore suggest a quasi-experimental approach to alleviate the problem of poor model fit.

It is necessary to isolate an explicit selection variable. This can often be done by arranging for the selection of all applicants whose test score is above some specified value. Applicants whose scores are below this value can also be selected as desirable, but for statistical calculations these applicants are not considered part of the selected group. This compromise arrangement should prove acceptable, administratively, in many situations.

A third issue concerns the extent to which selection has taken place. Lord and Novick [1968, p. 148] discourage the application of the correction formulas when extreme selection has occurred on the grounds that the accuracy of the formulas decrease as the selection ratio increases.

Caution must be used to ensure that formulas are not applied when extreme selection has taken place. It is quite clear that the accuracy of the formulas decreases as the selection ratio increases.

A fifth issue concerns the magnitude of the uncorrected validity coefficient. Greener and Osburn [1979] found corrected validities to be less accurate than uncorrected validities at the low range of the validity scale. Corrected correlation coefficients in general tend to be more accurate estimators than the uncorrected ones for moderate to large unrestricted population correlations [Greener & Osburn, 1979; Mendoza, Hart & Powell, 1991].

Campbell [1991, pp. 696-697], arguing from an estimation bias perspective, summarises the literature on the accuracy of the traditional Pearson-Thorndike restriction of range corrections as follows:

Based on the above literature, there seems little reason to avoid using corrections for restriction of range. Their shortcomings are not that they carry a large risk of overestimating the population value, but that they still yield large underestimates under a variety of realistic conditions.

Gross and Kagen [1983], and Greener and Osburn [1979], however, in contrast to Campbell [1991], maintain that corrections for restriction of range are not always advantageous. This apparent contradiction originates in differences in the operationalisation of the term accuracy. Instead of approaching accuracy from the perspective of the expected deviation of the [uncorrected or corrected] sample estimate from the actual population value [i.e.  $E[\hat{\rho}[X,Y] - \rho[X,Y]]$ ], Gross and Kagen [1983] argue in favour of the expected mean square error [EMSE] [i.e.  $E[\hat{\rho}[X,Y] - \rho[X,Y]]^2$ ] as a criterion of estimation accuracy. Gross and Kagen [1983, pp. 390-391] present the essence of their argument as follows:

It can be argued that EMSE represents a more meaningful criterion of the accuracy of estimation than the criterion of bias. The basis for the argument is that the EMSE criterion reflects not only the bias of an estimator, but also its sampling variance. .... More specifically for small sample sizes, and small  $\rho_{XY}$  values, the EMSE value for  $r_a$  can be substantially smaller than that for the corrected correlation [ $r_c$ ]. Thus, it can be advantageous to estimate the population correlation using the uncorrected correlation [ $r_a$ ] rather than the corrected correlation [ $r_c$ ].

The Gross and Kagen [1983] commentary should not be taken to mean that the Pearson-Thorndike corrections should not be applied. The conditions under which the uncorrected coefficient would result in a smaller EMSE are limited to a weak  $\rho[X,Y]$  relationship, extreme selection and small selected groups; the majority of situations would still favour the application of the correction formula.

#### 3.4 THE VALIDITY COEFFICIENT CORRECTED FOR THE JOINT ATTENUATING EFFECT OF EXPLICIT OR IMPLICIT RESTRICTION OF RANGE AND THE UNRELIABILITY OF THE CRITERION MEASUREMENTS

Although considerable literature exists regarding the correction of correlation coefficients for the separate attenuating effects of measurement error and restriction of range [Ghiselli, Campbell & Zedeck, 1981; Gulliksen, 1950; Held & Foley, 1994; Linn, 1983; Olson & Becker, 1983; Pearson, 1903; Ree, Carretta, Earles & Albert, 1994], relatively little attention has been given to the theory underlying the correction of a correlation coefficient for the joint effects of measurement error and restriction of



range [Bobko, 1983; Lee, Millar & Graham, 1982; Mendoza & Mumford, 1987; Schmidt, Hunter & Urry, 1976].

In a typical validation study, restriction of range and criterion unreliability are simultaneously present. Their effects combine to yield an attenuated validity coefficient that could severely underestimate the operational validity [Lee, Millar & Graham, 1982; Schmidt, Hunter & Urry, 1976]. It thus seems to make intuitive sense to double correct an obtained validity coefficient for the attenuating affect of both factors. The APA, however, through their Standards for Educational and Psychological Tests [APA, 1974, p.41], initially recommended that:

It is ordinarily unwise to make sequential corrections, as in applying a correction for attenuation to a coefficient already corrected for restriction of range. Chains of corrections may be useful in considering possible further research, but their results should not be seriously reported as estimates of population correlation coefficients.

Schmidt, Hunter and Urry [1976], though, consider the APA recommendation to be in error and propose that the obtained validity coefficient should be sequentially corrected for the effects of both restriction of range and criterion unreliability so as to obtain an estimate of the actual operational validity. The revised edition of the Standards for Educational and Psychological Tests [APA, 1985] subsequently also seems to have softened its position on this topic by abstaining from any comment. The stepwise correction procedure suggested by Schmidt, Hunter and Urry [1976] involves first correcting both the obtained validity and reliability coefficients for restriction of range since both coefficients apply only to a restricted applicant group and thus are to a greater or lesser extent negatively biased estimates of the operational reliability and validity coefficients. Equation 3.8 is suggested as an appropriate correction formula to correct the reliability coefficient for the attenuating affect of restriction of range if homogeneity of error variance across values of the range of true criterion scores can be assumed [i.e. the assumption is that applicants were selected in such a manner that the true score variance reduces whereas the error variance remains unaffected] [Feldt & Brennan, 1989; Ghiselli, Campbell & Zedeck, 1981; Guion, 1965; Gulliksen, 1950; Lee, Millar & Graham, 1982].

From the assumption of homogeneous error variance across values of the range of true criterion scores it follows that:

$$\sigma[y]\sqrt{(1 - \rho_{tty})} = \sigma[Y]\sqrt{(1 - \rho_{ttY})} \text{-----} 3.8.1$$

Squaring Equation 3.8.1:

$$\sigma^2[y](1 - \rho_{tty}) = \sigma^2[Y](1 - \rho_{ttY}) \text{-----} 3.8.2$$

Multiplying Equation 3.8.2 by  $1/\sigma^2[Y]$ :

$$(\sigma^2[y]/\sigma^2[Y])(1 - \rho_{tty}) = (1 - \rho_{ttY}) \text{-----} 3.8.3$$

Isolating the unrestricted reliability coefficient in Equation 3.8.3:

$$\rho_{ttY} = 1 - \{(\sigma[y]/\sigma[Y])^2(1 - \rho_{tty})\} \text{-----} 3.8$$

The assumption Equation 3.8 is based on, however, frequently does not hold [Feldt & Brennan, 1989]. A further problem with Equation 3.8 in the context of validation research, however, is that the criterion variance for the unrestricted group is logically impossible to obtain.

Schmidt, Hunter and Urry [1976] suggest an alternative expression [shown as Equation 3.9] which avoids the aforementioned problem.

$$\rho_{ttY} = [1 - [(1 - \rho_{tty})/(1 - \rho[x,y])(1 - (\sigma^2[X]/\sigma^2[x]))] \text{-----} 3.9$$

Depending on the nature of the selection and the variable for which both the restricted and unrestricted variance is known, the correction of the validity coefficient for the attenuating effect of restriction of range will proceed through the appropriate correction formula taken from Equations 3.3 - 3.6. The validity coefficient corrected for restriction of range will then subsequently be corrected for the attenuation effect of criterion unreliability by employing the results of the preceding first two steps [i.e. the reliability and validity coefficients corrected for restriction of range] in the traditional attenuation correction formula for the criterion only [Equation 3.2].

Lee, Miller and Graham [1982], however, point out that statistical and measurement theory permits a simpler two-step correction. According to the Lee, Miller and Graham [1982] approach the restricted criterion reliability coefficient is used to correct the restricted validity coefficient for the attenuating effect due to the unreliability of the criterion. This reliability attenuation-corrected validity coefficient is then subsequently corrected for the attenuating affect of restriction of range. The first step in the Schmidt, Hunter and Urry [1976] procedure is thus disposed of. Although the procedures suggested by Schmidt, Hunter and Urry [1976] and Lee, Miller and Graham [1982] seem to be conceptually distinct, Bobko [1983] points out that these two procedures are in fact arithmetically identical. Combining the two step-approach suggested by Lee, Miller and Graham [1982] into a single equation results in Equation 3.10 for the double-corrected validity coefficient [assuming Case 2 [Case A] selection produced the restriction of range] [Bobko, 1983].

$$\rho[X,TY] = \sigma[X]\rho[x,y]\rho[y,y]^{1/2} / \{\sigma^2[X]\rho^2[x,y]\rho[y,y]^{-1} + \sigma^2[x] - \sigma^2[x]\rho^2[x,y]\rho[y,y]^{-1}\}^{1/2} \text{-----} 3.10$$

Similar equations could be derived for the other possible conditions under which correlation estimation bias due to systematic selection could occur.

Mendoza & Mumford [1987] proposed a set of equations in terms of which correlation coefficients can be jointly corrected for:

- restriction of range directly on the predictor and unreliability in the predictor and the criterion; or
- restriction of range directly on the latent trait measured by the predictor and unreliability in the predictor and the criterion

Equation 3.11 shows the appropriate correction formula applicable when restriction of range occurs directly on the ability/latent trait measured by the predictor [Mendoza & Mumford, 1987]. The derivation of Equation 3.11 assumes a linear, homoscedastic regression of the criterion Y on the predictor X in the unrestricted population and in addition makes the two usual restriction of range assumptions that:

- the regression of actual job performance [i.e. the ultimate criterion] Y' on ability will not be affected by explicit selection on the latent trait represented by X'; and
- the ultimate criterion variance conditional on X' will not be altered by explicit selection on the latent trait measured by X [Mendoza & Mumford, 1987].

From the assumption that the regression of actual job performance [i.e. the ultimate criterion] Y' on ability will not be affected by explicit selection on the latent trait represented by X, it follows that:

$$\beta[t_y | t_x] = \beta[T_Y | T_X] \text{-----} 3.11.1$$

From the assumption that the ultimate criterion variance conditional on X' will not be altered by explicit selection on the latent trait measured by X, it follows that:

$$\sigma^2[t_y](1 - \rho^2[t_x t_y]) = \sigma^2[T_Y](1 - \rho^2[T_X, T_Y]) \text{-----} 3.11.2$$

However

$$\begin{aligned} \beta^2[t_y | t_x] &= \rho^2[t_y t_x](\sigma^2[t_y]/\sigma^2[t_x]) \\ &= \rho^2[t_y t_x]\{(\sigma^2[y]\rho_{ttx})/(\sigma^2[x]\rho_{ttx})\} \text{-----} 3.11.3 \end{aligned}$$

Similarly:

$$\beta^2[T_Y | T_X] = \rho^2[T_Y T_X](\sigma^2[T_Y]/\sigma^2[T_X])$$

$$= \rho^2[T_Y T_X] \{(\sigma^2[Y] \rho_{ttY}) / (\sigma^2[X] \rho_{ttX})\} \text{-----} 3.11.4$$

Substituting Equations 3.11.3 and 3.11.4 in Equation 3.11.1:

$$\rho^2[T_Y T_X] \{(\sigma^2[Y] \rho_{ttY}) / (\sigma^2[X] \rho_{ttX})\} = \rho^2[t_y t_x] \{(\sigma^2[y] \rho_{tty}) / (\sigma^2[x] \rho_{ttx})\} \text{-----} 3.11.5$$

Isolating the term  $\rho^2[T_Y T_X]$  in Equation 3.11.5 by multiplying by  $\{(\sigma^2[X] \rho_{ttX}) / (\sigma^2[Y] \rho_{ttY})\}$

$$\begin{aligned} \rho^2[T_Y T_X] &= \rho^2[t_y t_x] \{(\sigma^2[y] \rho_{tty}) / (\sigma^2[x] \rho_{ttx})\} \{(\sigma^2[X] \rho_{ttX}) / (\sigma^2[Y] \rho_{ttY})\} \\ &= \rho^2[t_y t_x] \{(\sigma^2[y] \rho_{tty} \sigma^2[X] \rho_{ttX}) / (\sigma^2[x] \rho_{ttx} \sigma^2[Y] \rho_{ttY})\} \text{-----} 3.11.6 \end{aligned}$$

However, according to Equation 3.1:

$$\rho^2[t_x t_y] = \rho^2[x, y] / (\rho_{ttX} \rho_{ttY}) \text{-----} 3.11.7$$

Substituting Equation 3.11.7 in Equation 3.11.6:

$$\begin{aligned} \rho^2[T_X, T_Y] &= \{\rho^2[x, y] / (\rho_{ttX} \rho_{ttY})\} \{(\sigma^2[y] \rho_{tty} \sigma^2[X] \rho_{ttX}) / (\sigma^2[x] \rho_{ttx} \sigma^2[Y] \rho_{ttY})\} \\ &= (\rho^2[x, y] \sigma^2[y] \sigma^2[X] \rho_{ttX}) / (\sigma^2[x] \rho^2_{ttX} \sigma^2[Y] \rho_{ttY}) \text{-----} 3.11 \end{aligned}$$

Equation 3.11 places rather formidable demands in as far as it requires the reliability and variance of both variables in both the restricted and unrestricted groups to be known. This seems to limit the practical value of Equation 3.11. If it is possible to calculate both  $\sigma^2[X]$  and  $\sigma^2[Y]$  [and not only one of the two], it seems more than probable that one would also be able to calculate  $\rho[X, Y]$ ,  $\rho_{ttX}$  and  $\rho_{ttY}$  and thus estimate  $\rho[T_X, T_Y]$  with the traditional attenuation correction formula [Equation 3.1]. The need to infer  $\rho[T_X, T_Y]$  indirectly via an equation like Equation 3.11, would then no longer exist. Mendoza and Mumford [1987] acknowledge the equation's requirement that the reliability of both measures be known in the restricted and unrestricted space, but do not regard this as a problem since the restricted and unrestricted reliabilities are related by Equation 3.8.

Equation 3.12 applies to the second, probably more prevalent, situation where restriction of range/selection occurs directly on the predictor [Mendoza & Mumford, 1987]. The derivation of Equation 3.12 assumes a linear, homoscedastic regression of the criterion Y on the predictor X in the unrestricted population and in addition makes the two usual restriction of range assumptions that:

- the regression of the criterion Y on the predictor will not be affected by explicit selection on the predictor X; and
- the criterion variance conditional on X will not be altered by explicit selection on X [Mendoza & Mumford, 1987].

From the assumption that the regression of the criterion Y on the predictor will not be affected by explicit selection on the predictor X, it follows that:

$$\beta[y|x] = \beta[Y|X] \text{-----} 3.12.1$$

From the assumption that the criterion variance conditional on X will not be altered by explicit selection on the predictor X, it follows that:

$$\sigma^2[y](1 - \rho^2[x,y]) = \sigma^2[Y](1 - \rho^2[X,Y]) \text{-----} 3.12.2$$

From Equation 3.12.1 it follows that:

$$\rho^2[x,y](\sigma^2[y]/\sigma^2[x]) = \rho^2[X,Y](\sigma^2[Y]/\sigma^2[X]) \text{-----} 3.12.3$$

Isolating the term  $\rho^2[X,Y]$  in Equation 3.12.3:

$$\rho^2[X,Y] = \rho^2[x,y](\sigma^2[y]\sigma^2[X]) / (\sigma^2[x]\sigma^2[Y]) \text{-----} 3.12.4$$

However, Equation 3.1 states that:

$$\rho[T_X, T_Y] = \rho[X,Y] / (\sqrt{\rho_{tt}X} \sqrt{\rho_{tt}Y})$$

Substituting Equation 3.12.4 in the square of Equation 3.1:

$$\rho^2[T_X, T_Y] = (\rho^2[x,y]\sigma^2[y]\sigma^2[X]) / (\sigma^2[x]\sigma^2[Y]\rho_{tt}X\rho_{tt}Y) \text{-----} 3.12.5$$

However,  $\sigma^2[Y]$  and  $\rho_{tt}Y$  probably would not be available.

Multiplying Equation 3.12.2 by  $1/\{\sigma^2[Y](1 - \rho^2[x,y])\}$ :

$$\sigma^2[y]/\sigma^2[Y] = (1 - \rho^2[X,Y]) / (1 - \rho^2[x,y]) \text{-----} 3.12.6$$

However, according to Equation 3.4:

$$\rho[X,Y] = (\sigma[X]/\sigma[x])\rho[x,y] / \{(\sigma^2[X]/\sigma^2[x])\rho^2[x,y] + 1 - \rho^2[x,y]\}^{1/2}$$

Squaring Equation 3.4:

$$\rho^2[X,Y] = (\sigma^2[X]/\sigma^2[x])\rho^2[x,y]/\{(\sigma^2[X]/\sigma^2[x])\rho^2[x,y] + 1 - \rho^2[x,y]\} \text{-----} 3.12.7$$

Let  $K$  represent  $(\sigma^2[X]/\sigma^2[x])$ . Equation 3.12.7 can then be rewritten as:

$$\rho^2[X,Y] = K\rho^2[x,y]/(K\rho^2[x,y] + 1 - \rho^2[x,y]) \text{-----} 3.12.8$$

Substituting Equation 3.12.8 in the numerator of Equation 3.12.6:

$$\begin{aligned} (1 - \rho^2[X,Y]) &= 1 - \{K\rho^2[x,y]/(K\rho^2[x,y] + 1 - \rho^2[x,y])\} \\ &= (K\rho^2[x,y] + 1 - \rho^2[x,y] - K\rho^2[x,y])/(K\rho^2[x,y] + 1 - \rho^2[x,y]) \\ &= (1 - \rho^2[x,y])/(K\rho^2[x,y] + 1 - \rho^2[x,y]) \text{-----} 3.12.9 \end{aligned}$$

Substituting Equation 3.12.9 in Equation 3.12.6:

$$\begin{aligned} \sigma^2[y]/\sigma^2[Y] &= (1 - \rho^2[X,Y])/(1 - \rho^2[x,y]) \\ &= (1 - \rho^2[x,y])/(K\rho^2[x,y] + 1 - \rho^2[x,y])/(1 - \rho^2[x,y]) \\ &= (K\rho^2[x,y] + 1 - \rho^2[x,y])^{-1} \text{-----} 3.12.10 \end{aligned}$$

Substituting Equation 3.12.10 in Equation 3.12.5:

$$\begin{aligned} \rho^2[T_X, T_Y] &= (\rho^2[x,y]\sigma^2[y]\sigma^2[X])/(\sigma^2[x]\sigma^2[Y]\rho_{ttX}\rho_{ttY}) \\ &= (\rho^2[x,y])(\sigma^2[y]/\sigma^2[Y])(\sigma^2[X]/\sigma^2[x])(1/\rho_{ttX})(1/\rho_{ttY}) \\ &= (\rho^2[x,y])K(1/\rho_{ttX})(1/\rho_{ttY})(\sigma^2[y]/\sigma^2[Y]) \\ &= \{(\rho^2[x,y])K\}/\{(\rho_{ttX}\rho_{ttY})(K\rho^2[x,y] + 1 - \rho^2[x,y])\} \text{-----} 3.12.11 \end{aligned}$$

However, the problem of the unavailability of  $\rho_{ttY}$  still exists.

Substituting Equation 3.12.10 in Equation 3.8:

$$\begin{aligned} \rho_{ttY} &= 1 - (\sigma^2[y]/\sigma^2[Y])(1 - \rho_{tty}) \\ &= 1 - \{(K\rho^2[x,y] + 1 - \rho^2[x,y])^{-1}\}(1 - \rho_{tty}) \text{-----} 3.12.12 \end{aligned}$$

Therefore:

$$\begin{aligned} \rho_{ttY} &= \{(K\rho^2[x,y] + 1 - \rho^2[x,y]) - 1 + \rho_{tty}\}/(K\rho^2[x,y] + 1 - \rho^2[x,y]) \\ &= (K\rho^2[x,y] + 1 - \rho^2[x,y] - 1 + \rho_{tty})/(K\rho^2[x,y] + 1 - \rho^2[x,y]) \\ &= (K\rho^2[x,y] - \rho^2[x,y] + \rho_{tty})/(K\rho^2[x,y] + 1 - \rho^2[x,y]) \text{-----} 3.12.13 \end{aligned}$$



Substituting Equation 3.12.13 in Equation 3.12.11:

$$\begin{aligned}\rho^2[T_X, T_Y] &= (\rho^2[x, y]K) / \{(\rho_{ttX}\rho_{ttY})(K\rho^2[x, y] + 1 - \rho^2[x, y])\} \\ &= (\rho^2[x, y])K / \{(\rho_{ttX}(K\rho^2[x, y] - \rho^2[x, y]) + \\ &\quad \rho_{ttY}) / (K\rho^2[x, y] + 1 - \rho^2[x, y])(K\rho^2[x, y] + 1 - \rho^2[x, y])\} \\ &= (\rho^2[x, y])K / \{(\rho_{ttX}(K\rho^2[x, y] - \rho^2[x, y]) + \rho_{ttY})\} \text{-----} 3.12\end{aligned}$$

Equation 3.12, however, still has rather limited utility in applied validation research. Its primary deficiency lies in the fact that it also corrects the correlation coefficient for the unreliability of predictor measurements. Correcting for unreliability in the predictor in a validation context is misleading. It would be of relative little value to know the validity of a perfectly reliable predictor when such an infallible measuring instrument can never be available for operational use [Lee, Miller & Graham, 1982; Nunnally, 1978; Schmidt, Hunter & Urry, 1976;]. This problem can, however, relatively easily be rectified [Schepers, 1996] as shown by Equation 3.13.

From the assumption that the regression of the criterion Y on the predictor will not be affected by explicit selection on the predictor X, it follows that:

$$\beta[y|x] = \beta[Y|X] \text{-----} 3.13.1$$

From the assumption that the criterion variance conditional on X will not be altered by explicit selection on the predictor X, it follows that:

$$\sigma^2[y](1 - \rho^2[x, y]) = \sigma^2[Y](1 - \rho^2[X, Y]) \text{-----} 3.13.2$$

Equation 3.13.1 implies that:

$$\rho^2[x, y](\sigma^2[y]/\sigma^2[x]) = \rho^2[X, Y](\sigma^2[Y]/\sigma^2[X]) \text{-----} 3.13.3$$

Isolating the term  $\rho^2[X, Y]$  in Equation 3.13.3:

$$\rho^2[XY] = (\rho^2[x, y]\sigma^2[y]\sigma^2[X]) / (\sigma^2[x]\sigma^2[Y]) \text{-----} 3.13.4$$

However, according to Equation 3.1:

$$\rho[X, T_Y] = \rho[X, Y] / \sqrt{\rho_{ttY}}$$

Squaring Equation 3.1 and substituting Equation 3.13.4:

$$\rho^2[X, T_Y] = (\rho^2[x, y]\sigma^2[y]\sigma^2[X]) / (\sigma^2[x]\sigma^2[Y]\rho_{tt}Y) \text{-----} 3.13.5$$

However,  $\sigma^2[Y]$  and  $\rho_{tt}Y$  probably would not be available.

Multiplying Equation with  $1/\{\sigma^2[Y](1 - \rho^2[x, y])\}$ :

$$\sigma^2[y]/\sigma^2[Y] = (1 - \rho^2[X, Y]) / (1 - \rho^2[x, y]) \text{-----} 3.13.6$$

However, according to Equation 3.4:

$$\rho[X, Y] = (\sigma[X]/\sigma[x])\rho[x, y] / \{(\sigma^2[X]/\sigma^2[x])\rho^2[x, y] + 1 - \rho^2[x, y]\}^{1/2}$$

Let K denote  $(\sigma^2[X]/\sigma^2[x])$ .

Squaring Equation 3.4 and substituting K:

$$\begin{aligned} \rho^2[X, Y] &= \{(\sigma^2[X]/\sigma^2[x])\rho[x, y]\} / \{(\sigma^2[X]/\sigma^2[x])\rho^2[x, y] + 1 - \rho^2[x, y]\} \\ &= K\rho[x, y] / (K\rho^2[x, y] + 1 - \rho^2[x, y]) \text{-----} 3.13.7 \end{aligned}$$

Substituting Equation 3.13.7 in the numerator of Equation 3.13.6:

$$\begin{aligned} (1 - \rho^2[X, Y]) &= 1 - \{K\rho[x, y] / (K\rho^2[x, y] + 1 - \rho^2[x, y])\} \\ &= (K\rho^2[x, y] + 1 - \rho^2[x, y] - K\rho[x, y]) / (K\rho^2[x, y] + 1 - \rho^2[x, y]) \\ &= (1 - \rho^2[x, y]) / (K\rho^2[x, y] + 1 - \rho^2[x, y]) \text{-----} 3.13.8 \end{aligned}$$

Substituting Equation 3.13.8 in Equation 3.13.6:

$$\begin{aligned} \sigma^2[y]/\sigma^2[Y] &= (1 - \rho^2[X, Y]) / (1 - \rho^2[x, y]) \\ &= (1 - \rho^2[x, y]) / \{(K\rho^2[x, y] + 1 - \rho^2[x, y])(1 - \rho^2[x, y])\} \\ &= (K\rho^2[x, y] + 1 - \rho^2[x, y])^{-1} \text{-----} 3.13.9 \end{aligned}$$

Substituting Equation 3.13.9 in Equation 3.13.5:

$$\begin{aligned} \rho^2[X, T_Y] &= (\rho^2[x, y]\sigma^2[y]\sigma^2[X]) / (\sigma^2[x]\sigma^2[Y]\rho_{tt}Y) \\ &= (\rho^2[x, y])(\sigma^2[y]/\sigma^2[Y])(\sigma^2[X]/\sigma^2[x])(1/\rho_{tt}Y) \\ &= (\rho^2[x, y])K(1/\rho_{tt}Y)(\sigma^2[y]/\sigma^2[Y]) \\ &= (\rho^2[x, y]K) / \{(\rho_{tt}Y)(K\rho^2[x, y] + 1 - \rho^2[x, y])\} \text{-----} 3.13.10 \end{aligned}$$

However,  $\rho_{ttY}$  probably would not be known.

Substituting Equation 3.13.9 in Equation 3.8:

$$\begin{aligned}\rho_{ttY} &= 1 - (\sigma^2[y]/\sigma^2[Y])(1 - \rho_{tty}) \\ &= 1 - (K\rho^2[x,y] + 1 - \rho^2[x,y])^{-1}(1 - \rho_{tty})\end{aligned}\quad 3.13.11$$

Therefore:

$$\begin{aligned}\rho_{ttY} &= \{(K\rho^2[x,y] + 1 - \rho^2[x,y]) - 1 + \rho_{tty}\} / (K\rho^2[x,y] + 1 - \rho^2[x,y]) \\ &= (K\rho^2[x,y] + 1 - \rho^2[x,y] - 1 + \rho_{tty}) / (K\rho^2[x,y] + 1 - \rho^2[x,y]) \\ &= (K\rho^2[x,y] - \rho^2[x,y] + \rho_{tty}) / (K\rho^2[x,y] + 1 - \rho^2[x,y])\end{aligned}\quad 3.13.12$$

Substituting Equation 3.13.12 in Equation 3.13.10:

$$\begin{aligned}\rho^2[X, TY] &= (\rho^2[x,y]K) / \{\rho_{ttY}(K\rho^2[x,y] + 1 - \rho^2[x,y])\} \\ &= (\rho^2[x,y]K) / \{(K\rho^2[x,y] - \rho^2[x,y] + \rho_{tty}) / (K\rho^2[x,y] + 1 - \rho^2[x,y])\} \\ &= \rho^2[x,y]K / (K\rho^2[x,y] - \rho^2[x,y] + \rho_{tty})\end{aligned}\quad 3.13.13$$

Write:

$$\rho[X, TY] = \sqrt{\{\rho^2[x,y]K / (K\rho^2[x,y] - \rho^2[x,y] + \rho_{tty})\}}\quad 3.13$$

Equation 32 provides a joint correction of the correlation/validity coefficient for restriction of range directly on the predictor and the unreliability of the criterion. Multiplying the denominator and numerator of Equation 32 by  $\sigma[x]/\sqrt{\rho_{tty}}$ , it can be shown that Equation 32 is in fact identical to Equation 5 presented by Bobko [1983] based on the two-step procedure suggested by Lee, Miller and Graham [1982]. A hitherto unrecognised agreement between the work of Bobko [1983] and Mendoza and Mumford [1987] on the joint correction of the correlation/validity coefficient is therefore established. The correction formula derived from the work by the Mendoza and Mumford [1987], furthermore, is computationally slightly less cumbersome than the formula suggested by Bobko [1983].

How does Equation 3.13 affect the magnitude of the validity coefficient? The reaction of the double corrected correlation coefficient to changes in  $K = \phi$ , the reliability coefficient and the attenuated correlation coefficient, is graphically illustrated in Figures 3.13 - 3.16. The validity coefficient jointly corrected for Case B restriction of range and criterion unreliability was mapped onto a surface defined by  $0.05 \leq \rho[x,y] \leq 0.90$ ,  $0.10 \leq \rho_{tty} \leq 0.9$  and  $1 \leq K \leq 4$  through a SAS program feeding a selection of surface

coordinates into Equation 3.13. Figures 3.13 - 3.16 indicate that the amount of benefit derived from Equation 3.13 increases as  $K$  increases and  $\rho_{TTY}$  decreases. The uncorrected validity coefficient  $\rho[x,y]$  [i.e. the observed validity coefficient for the attenuating effect of both restriction of range and criterion unreliability] provides a too conservative description of the actual correlation existing between  $X$  and  $T_Y$ . The extent to which  $\rho[x,y]$  underestimate  $\rho[X,T_Y]$  increases as the restriction of range becomes more severe and the reliability of the criterion scores declines. The corrected validity coefficient  $\rho[X,T_Y]$  seems to be a positive curvilinear function of  $\rho[x,y]$ , with the degree of curvilinearity diminishing as the attenuated validity coefficient increases. The corrected validity coefficient, similarly, increases curvilinearly with an increase in the attenuated validity coefficient, with the degree of curvilinearity increasing as  $K = \sigma^2[X]/\sigma^2[x]$  increases. Relatively more, therefore, is gained by correcting an attenuated validity coefficient observed in the lower region of the validity scale than in the upper region of the scale.

The findings reported here clearly indicates the dramatic consequence of correcting the observed validity coefficient for the attenuating effect of both restriction of range and criterion unreliability, especially when severe range restriction occurred and the criterion measures suffer from low reliability. Not to correct the observed validity coefficient will severely underestimate the actual validity of the selection procedure for the applicant population. Lee, Miller and Graham [1982], and Bobko [1983] concur that all the available evidence argue in favor of jointly correcting the validity coefficient for the attenuating effect of both range restriction and the unreliability of the criterion. Lee, Miller and Graham [1982] found most corrected validity coefficients to be slight overestimates of the true validity coefficient. In direct contrast to the findings reported by Lee, Miller and Graham [1982], Bobko [1983] concludes that, on average, the double corrected validity coefficient will still underestimate the operational validity coefficient. The research reported here does not permit any comment on bias in the corrected validity coefficient.

A further, less serious, limitation of both Equations 3.13 and 3.12 concerns the premise that selection can only occur directly on the predictor. Case C conditions [indirect restriction of range on the predictor and the criterion through direct selection on a third variable] probably constitute the predominant environment in which restriction of range corrections are required. Again, however, this problem can relatively easily be rectified by substituting the Case B restriction of range correction formula in the derivation of Equation 3.13 and Equation 3.12 with the appropriate Case C correction formula [Gulliksen, 1950; Thorndike, 1949;].

Ptty

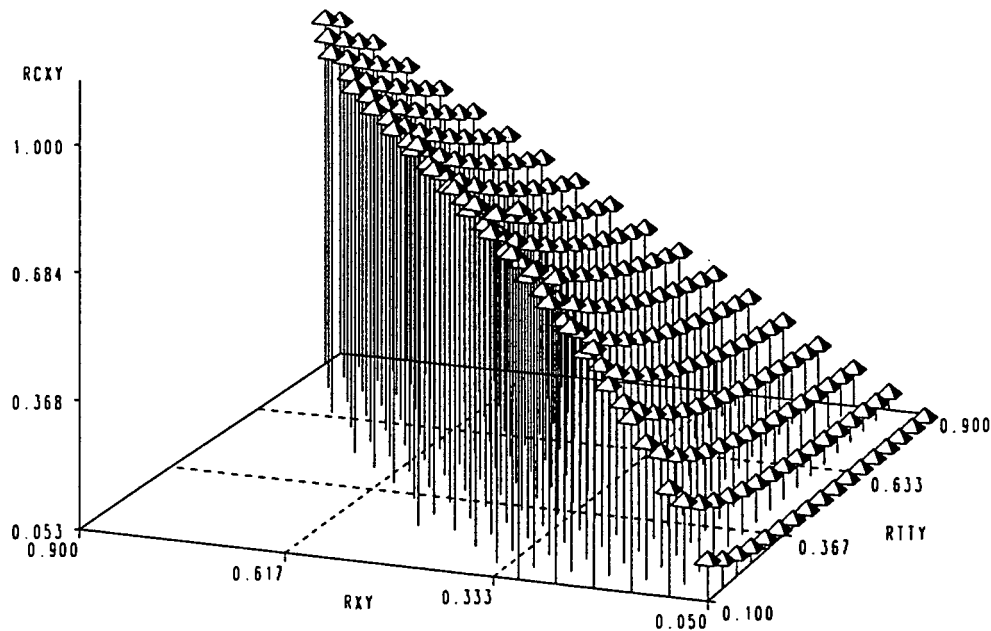


Figure 3.13: The validity coefficient jointly corrected for Case 2 [Case A] restriction of range and criterion unreliability as a function of  $\rho_{TTY}$  and  $\rho_{[x,y]}$  for  $K$  fixed at 1

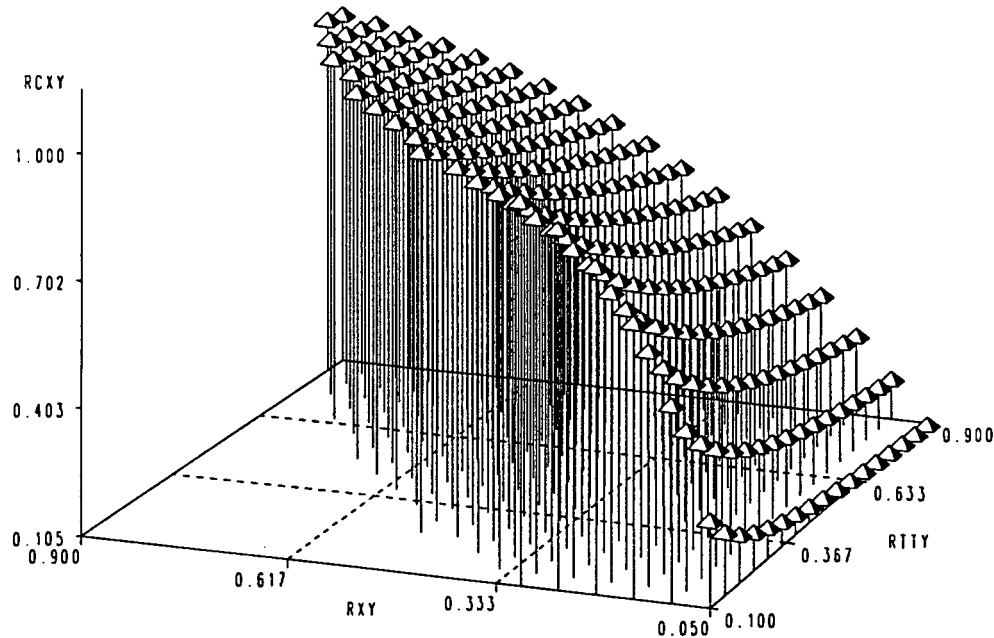


Figure 3.14: The validity coefficient jointly corrected for Case 2 [Case A] restriction of range and criterion unreliability as a function of  $\rho_{TTY}$  and  $\rho_{[x,y]}$  for  $K$  fixed at 2

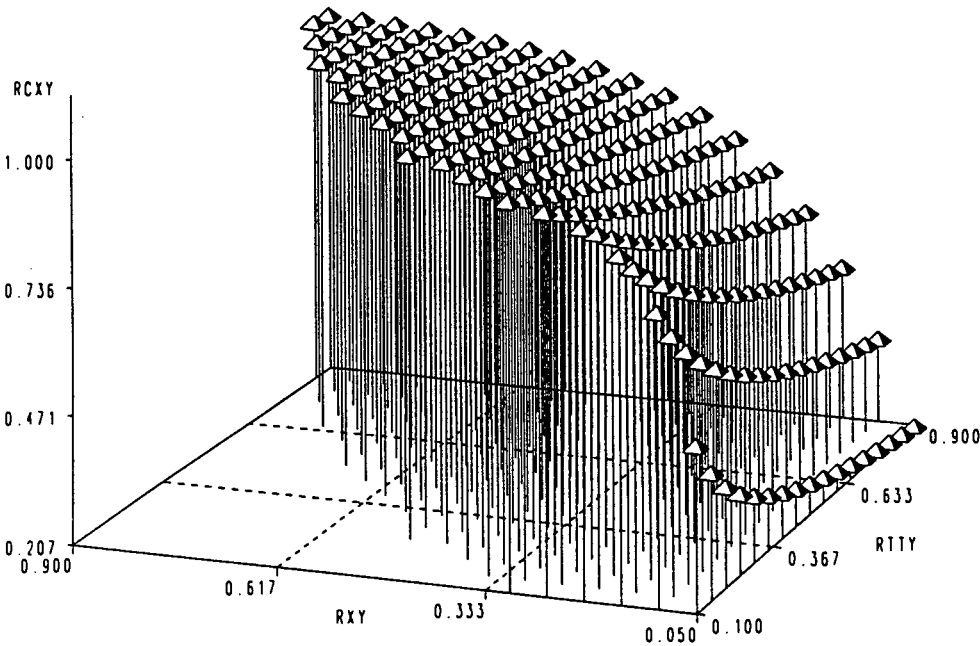


Figure 3.15: The validity coefficient jointly corrected for Case 2 [Case A] restriction of range and criterion unreliability as a function of  $\rho_{tty}$  and  $\rho[x,y]$  for  $K$  fixed at 4

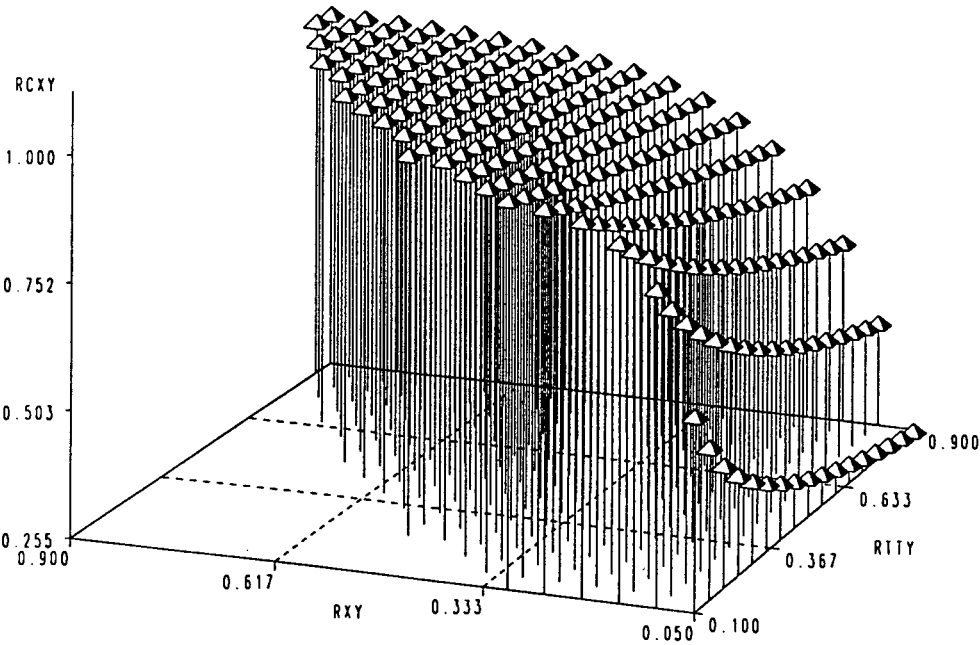


Figure 3.16: The validity coefficient jointly corrected for Case 2 [Case A] restriction of range and criterion unreliability as a function of  $\rho_{tty}$  and  $\rho[x,y]$  for  $K$  fixed at 5



ΞΨζξψΖξψζξψΖξψζξψΖξψζξψΖ

## CHAPTER 4

### STATISTICAL INFERENCE REGARDING CORRELATIONS CORRECTED FOR THE SEPARATE AND COMBINED ATTENUATING EFFECT OF RESTRICTION OF RANGE AND CRITERION UNRELIABILITY

The purpose of the following chapter is firstly to provide a brief description of the place and role of statistical inference in validation research. The function of sampling distributions of statistics in statistical inference will subsequently be sketched. The different procedures available to establish the necessary knowledge on the statistical properties of sampling distributions of statistics will thereafter be discussed. A survey of the existing literature on the characteristics of the sampling distributions of correlations corrected for the attenuating effect of predictor and criterion unreliability, criterion unreliability, Case 1 [Case B], Case 2 [Case A] and Case 3 [Case C] restriction of range and the combined effect of criterion unreliability and Case 2 [Case A] restriction of range will be presented. The behaviour of the standard error of the corrected correlation coefficient, in isolation and in relation to the standard error of the uncorrected coefficient, will finally be examined by adjusting the levels of relevant parameters and portraying the reaction visually in three dimensional graphic representations.

#### 4.1 INFERENTIAL STATISTICS IN VALIDATION RESEARCH

Due to practical considerations, the investigation of research questions are normally conducted on samples drawn/obtained from larger populations. Although forced by practical considerations to investigate only a sub-section of the population, the focus of interest, however, still is on the [unknown] value of one or more specific parameters  $[\theta_i]$  characterising the population and not the values of the corresponding statistics  $[\theta^*_i]$  characterising the sample. The statistics  $[\theta^*_i]$  are normally considered relevant only because they reflect, albeit with less than perfect accuracy, the parameter of interest [assuming the sample to be representative of the population of interest]. The inability of sample statistics/estimators to precisely mirror the corresponding parameter in the parent population lies in the inherent incapability of a sample to represent/substitute its source population perfectly. The researcher thus has to generalise beyond the obtained sample results under conditions of uncertainty. Three basic inferential options are available:

- obtain a point estimate  $[\theta^*]$  of the unknown population parameter  $[\theta]$ ; and/or
- obtain a confidence interval estimate of the unknown population parameter  $[\theta]_{lo} \geq \theta \geq \theta_{up}$ ; and/or
- decide on the tenability of a hypothesis  $[\theta = \theta_0]$  on the value of the unknown population parameter by estimating the probability of the sample statistic

conditional on the hypothesised value of the unknown population parameter  $[P[\theta^{\wedge}=\theta^{\wedge}_1 | \theta=\theta_0]]$ .

Assessment of the accuracy of point estimates would be possible only if the sampling distribution of the estimator is known or could be simulated. Similarly, the latter two options would be viable only if the "behavior" of the sample estimator were known [i.e. could be statistically and mathematically described] or could be simulated. Thus, either the distributional form of the sampling distribution of the estimator must be known or an array of possible values for the sample estimator must be generated through a series of independent samples. The distributional form of the sampling distribution of the estimator is considered known if a mathematical model/probability density function can be specified that describes the sampling distribution of the estimator exactly or by approximation.

All correction formula introduced in chapter 3 [and all other statistical expressions in the preceding chapters] were presented as population parameter and thus are, in terms of the preceding argument, normally not directly calculable. It is, however, possible to estimate the parameter of interest by substituting the parameters contained in the formula with the appropriate sample estimators. This strategy rests on the critical assumption that the statistics obtained in an applied research situation may be considered valid estimates of the respective parameters in either the unrestricted [i.e. entire] or selected applicant populations. This in turn rests on the equally critical presumption that the unrestricted and restricted samples of actual observations may be considered representative of the two respective applicant populations. This in effect means that the actual applicant groups studied must be equivalent to random samples drawn from the theoretical populations of interest. Too often, however, there is sufficient reason to believe that this is in fact not the case [Mendoza & Mumford, 1987].

Even if these assumption would be warranted, the corrected sample correlation coefficient  $[\rho^{\wedge}_c]$  would, however, still only reflect with less than perfect accuracy, the parameter of interest  $[\rho_c]$  due to sampling error. Consequently a need for some indication on the accuracy of  $\rho^{\wedge}_c$  as a point estimate of the unattenuated parameter  $\rho_c$  arises. The construction of a confidence interval estimate of the unknown corrected population parameter  $[\rho_u \geq \rho_c \geq \rho_l]$  would, furthermore, provide additional formal acknowledgement of the inherent uncertainty in generalising the corrected sample correlation coefficient beyond the confines of the validation sample. Furthermore, due to the inherent incapability of any validation sample to represent/substitute its source applicant population perfectly, the possibility always exists that the corrected sample correlation coefficient is attributable purely to chance [i.e. the possibility always exists that  $\rho_c = 0$  even though  $\rho^{\wedge}_c \neq 0$ ]. The unassailability of a defense of a selection procedure, based on validity evidence derived from a validation sample, thus clearly depends [amongst other things] on the magnitude of the probability of the sample statistic  $[\rho^{\wedge}_c]$  conditional on the hypothesised position that the unknown population parameter equals zero [i.e.  $P[\rho^{\wedge}_c=\rho^{\wedge}_{c1} | \rho=0]$ ]. Should the conditional probability not be sufficiently small [i.e. smaller than at least 0,05], the credibility

of the evidence lead in defense of the selection procedure will necessarily suffer. Consequently the need also exists to estimate  $\alpha_b = P[\hat{\rho}_c = \hat{\rho}_{c1} | \rho_c = 0]$ .

The feasibility of actually implementing any one or more of the aforementioned three options would, however, depend on the familiarity with the sampling distribution of the corrected correlation coefficient.

#### 4.2. THE NEED FOR INFORMATION ON THE SAMPLING DISTRIBUTION OF THE VALIDITY COEFFICIENT CORRECTED FOR THE ATTENUATING EFFECT OF THE UNRELIABILITY OF MEASUREMENTS AND/OR RESTRICTION OF RANGE

The sampling distribution of the Pearson product moment correlation coefficient under different sampling conditions has been well documented [Guilford & Fruchter, 1978; Hays, 1971]. The construction of confidence intervals around point estimates and the testing of hypotheses regarding population parameters consequently are relatively easy to perform. The traditional sampling theory, however, is not appropriate for testing hypotheses on the corrected validity coefficient, constructing confidence intervals around the corrected estimate or assessing the accuracy of point estimates of the corrected parameter [Forsyth, 1971; Mendoza, Hart & Powell, 1991]. Not knowing the sampling distribution applicable to such corrected estimates would seriously impede the ability of the validation researcher to build a convincing argument in defense of a selection procedure. There would be simply no way of rationally deciding whether the obtained sample results are attributable to chance or not [Cook, Campbell & Peracchio, 1991].

Logically the same problem experienced with testing hypothesis with regards to the corrected validity coefficient could potentially also apply to any other statistic of relevance to the argument lead in defense of a selection procedure [e.g. the regression coefficients defining the selection decision function], should such statistics also be affected by restriction of range and/or criterion unreliability or should the sampling variance of such statistics be unknown [e.g. the success ratio,  $S_v$ ].

The Standards for Educational and Psychological Testing [American Psychological Association, American Educational Research Association & National Council on Measurement in Education, 1974, p. 41] put forward the position that hypothesis on the statistical significance of corrected validity coefficients should not be tested:



Where correlation coefficients are corrected for attenuation or restriction of range, full information relevant to the correction should be presented. If such corrections are made, significance tests should be made with the uncorrected correlation coefficients.

If the implicit argument is that the sampling distribution of the corrected correlation coefficient differs from that of the uncorrected coefficient at least in terms of dispersion and thus that the normal test statistic do not apply to significance tests, the statement is accurate. The statement is, however, misleading in as far as it is implied that corrected correlation coefficients should and/or could under no circumstances be subjected to significance tests.

Moran [1970] indicates that hypotheses on functions of the Pearson correlation coefficient of the form  $H_0: f(\rho[X,Y]) = 0$  may be tested by expressing the obtained sample estimate in terms of standard error of the function of the correlation. Moran [1970] contends that  $f(\hat{\rho}[X,Y])/\sqrt{\text{Var}(f(r[X,Y]))}$  has, asymptotically, a standard normal [0,1] distribution. It thus follows that significance tests on corrected correlation/validity coefficients are possible if, but only if, an appropriate expression [or value] for the standard error of the corrected coefficient can be derived. Similarly, the construction of standard normal confidence intervals for corrected correlation coefficients should be a viable enterprise, provided appropriate standard error expressions [or values] are available [Bobko & Rieck, 1980].

#### 4.3 DIFFERENT PROCEDURES FOR INVESTIGATING THE SAMPLING DISTRIBUTIONS OF STATISTICAL ESTIMATORS

Two fundamental approaches exist in terms of which the sampling distribution of an estimator can be investigated. The one approach requires no assumptions about the form of the underlying population distribution and hence is referred to as a distribution free approach. The second approach, in contrast, starts by explicitly specifying the [actual or assumed] population distribution of the dependent variable and consequently is known as a non-distribution free approach. The terms non-parametric and parametric approaches are sometimes wrongfully used to refer to the aforementioned two approaches. In the case of the correlation coefficient both approaches are, by definition, parametric in nature.

#### 4.4 NON-DISTRIBUTION FREE ESTIMATES OF STATISTICAL ERROR

Two non-distribution free alternatives exist in terms of which the sampling distribution of an estimator could be examined:

- an experimental/empirical method; or
- an analytical/mathematical method.

The first method uses Monte Carlo resampling from a population with a known distribution to observe the behaviour of the sample estimator and to infer the statistical characteristics of the sampling distribution. The second method, in contrast, derives expressions for the statistical characteristics of the sampling behaviour of a sample estimator through mathematical analysis from explicit population distribution assumptions.

Although both methods seem suited to examine the sampling distribution of the corrected correlation coefficient under specific known parametric conditions, the analytical method seems to dominate more recent non-distribution free investigations. Kelley [1947] distinguishes the following analytic methods in terms of which the standard error of a statistic can be derived, if the statistic is a function of a statistic with a known standard error:

- the binomial expansion method;
- substitution of statistical deviations for calculus differentials method;
- expanding by a Taylor series method or Delta method; and
- logarithmic differentials method.

The Delta method [Rao, 1973] seems to be the preferred analytical approach.

#### 4.4.1 Delta Method

Assume an estimator which is a function of the Pearson product-moment correlation  $\rho^*[X,Y]$ . Let such an estimator be denoted by  $f(\rho^*[X,Y])$ . Using the Taylor expansion Kendall and Stuart (1977, p. 247) derive the expression shown as Equation 4.1 for the variance of  $f(\rho^*[X,Y])$ .

$$\sigma^2\{f(\rho^*[X,Y])\} = \Sigma[f'_i(\rho)]^2 \text{Var}[r_i] + \Sigma f'_i(\rho)f'_j(\rho) \text{Cov}[r_i, r_j] + [\text{terms of order } n^{-2} \text{ or less}] \text{-----} 4.1$$

Expressions for the asymptotic variance of  $\rho^*[X,Y]$  and the asymptotic covariance between two correlations computed on the same sample are required to execute Equation 4.1 [Bobko & Rieck, 1980]. The asymptotic variance of the correlation coefficient obtained from a bivariate normal population with parameter  $\rho[X,Y]$ , is given by Equation 4.2 [Kendall & Stuart, 1977, p. 251].

$$\sigma^2(\rho^*[X,Y]) = (1-\rho^2[X,Y])^2/n \text{-----} 4.2$$



Assuming multivariate normality, the asymptotic covariance between two correlations computed on the same sample is shown as Equation 4.3 [Kelley, 1947, p. 553] for the case where one variable is common to both correlations.

$$\sigma[r[X,Y],r[X,Z]]=\{\rho[X,Y](1-\rho^2[Y,Z]-\rho^2[X,Z])^{-1/2}(\rho[Y,Z]\rho[X,Z])(1-\rho^2[Y,Z]-\rho^2[X,Y]-\rho^2[X,Z])\}/n \text{-----} 4.3$$

## 4.5 DISTRIBUTION FREE ESTIMATES OF STATISTICAL ERROR

The trouble is that analytically or empirically derived standard errors for estimators more complex than the traditional descriptive statistics often do not exist. No mathematical expression thus exists that captures the capricious behaviour of the statistical estimator across repeated samples from the same parent population. Furthermore, the parametric assumptions underlying the analytically derived standard error expressions are often not satisfied, thus raising suspicion on the applicability of the expressions. Distribution free resampling techniques represent a feasible solution to these problems. A Monte Carlo approach would, however, normally not present a feasible solution in an applied context since information on the population distribution and its parameters are normally not available in applied validation research.

The bootstrap and jackknife are the two most popular data-resampling techniques used in inferential statistical analysis [Shao & Tu, 1995].

### 4.5.1 Bootstrap Estimates Of Statistical Error

The bootstrap represents an approach to the investigation of the sampling behavior of sample estimators which:

- relieves the analyst from having to make any parametric assumptions about the form of the underlying population distribution; and
- enables the analyst to proceed with inferential analysis even if no mathematical expressions on the sampling behaviour of the estimator exist.

The term bootstrap is derived from the expression to pull yourself up by your own bootstraps generally believed to originate from one of Rudolph Raspe's Adventures of Baron von Münchhausen [Efron & Tibshirani, 1993]. The term appears to be appropriate since the bootstrap procedure in essence

represents a seemingly impossible attempt to simulate the behaviour of a sample statistic across a large number of independent samples taken from a single parent population from data available only in a single sample [Diaconis & Efron, 1983; Efron, 1982; Efron & Tibshirani, 1993; Lunneborg, 1985; Shao & Tu, 1995].

Assume a population  $\Pi$  of potential observations characterised by the parameter  $\theta$ . Assume further a random sample  $\Psi$  of  $n$  observations drawn independently from population  $\Pi$ . Should the statistic  $\theta^{\wedge}$  corresponding to the parameter  $\theta$  in  $\Pi$  be calculated, an estimate of  $\theta$  would be obtained. A succession of  $m$  [bootstrap] samples  $\Psi_{bi}$  of size  $n$  are subsequently drawn randomly with replacement from the original sample  $\Psi$ . From each bootstrap sample  $\Psi_{bi}$ , a bootstrap estimate  $\theta^{\wedge}_{bi}$  is obtained. The fundamental bootstrap proposition is that the distribution of the bootstrap estimates  $\theta^{\wedge}_{bi}$  [i.e. the sampling distribution of  $\theta^{\wedge}_b$ ] will provide a sufficient approximation of the [true] sampling distribution derived analytically/theoretically from specific parametric assumptions and/or derived empirically through the classical Monte Carlo generation of  $m$  independent random samples  $\Psi_i$  of size  $n$  from  $\Pi$  [Diaconis & Efron, 1983; Efron, 1982; Efron & Tibshirani, 1993; Lunneborg, 1985; Shao & Tu, 1995]. From this conjecture follows that as the number of bootstrap samples approach infinity, the standard deviation of the distribution of bootstrap estimates, shown as Equation 4.4, approaches the standard error of the estimator  $\theta^{\wedge}$ .

$$\sigma[\theta^{\wedge}] \approx \sigma[\theta^{\wedge}_b] = \{\sum(\theta^{\wedge}_{bi} - (\sum \theta^{\wedge}_{bi}/m))^2 / (m-1)\}^{1/2} \text{-----} 4.4$$

#### 4.5.1.1 Bootstrap Confidence Intervals

Once the standard error of the estimator is obtained from the distribution of the bootstrap estimates  $\theta^{\wedge}_{bi}$  [i.e. the standard deviation of the bootstrap distribution], a parametric bootstrap confidence interval around the obtained sample estimate  $\theta^{\wedge}$  can be computed under the assumption that the sampling distribution of  $\theta^{\wedge}_b$  sufficiently corresponds to a familiar mathematical model [e.g. Gauss or Student distributions]. Assuming the sampling distribution to be approximately standard normal, a parametric confidence interval for the unknown parameter  $\theta$  can be obtained, with coverage probability equal to  $[1-\alpha]$ , through Equation 4.5 [Efron & Tibshirani, 1986; Efron & Tibshirani, 1993; Thompson, 1991].

$$\theta \in \theta^{\wedge} \pm \sigma^{\wedge}[\theta^{\wedge}_b] z_{\alpha/2} \text{-----} 4.5$$

Efron & Tibshirani [1993] suggest that the standard parametric confidence interval can be improved for small samples by substituting the standard normal critical values [given  $\alpha$ ] with critical values derived from the Student t distribution at appropriate degrees of freedom.

The bootstrap-t confidence interval differs from the aforementioned approaches in as far as it does not require apriori distributional assumptions. The bootstrap-t method requires the generation of  $m$  bootstrap samples. For each bootstrap sample the bootstrap estimate  $\theta^*_i$  is standardised by deviating it from the initial sample estimate  $\theta^*$  and expressing the deviation in terms of the estimated standard error of the bootstrap estimate [i.e.  $Z = (\theta^*_i - \theta^*) / \sigma^*[\theta^*_i]$ ]. The estimation of the standard error of  $\theta^*_i$  would require a further  $g$  bootstrap samples drawn with replacement from the  $i^{\text{th}}$  bootstrap sample; hence a potentially formidable total number of  $mg$  bootstrap samples are required to construct a bootstrap-t confidence interval [Shao & Tu, 1995]. The bootstrap-t confidence interval is subsequently constructed by "reading off" the Z-values corresponding to the  $100[\alpha/2]^{\text{th}}$  and  $100[1-\alpha/2]^{\text{th}}$  percentiles and inserting them with the sample estimate  $\theta^*$  and the bootstrap estimate of the standard error into Equation 4.6 [Efron & Tibshirani, 1993].

$$\theta \in \theta^* \pm \sigma^*[\theta^*_i] Z_{\alpha/2} \text{-----} 4.6$$

Whereas the standard normal and Student-t bootstrap confidence intervals are symmetric around the point estimate  $\theta^*$ , the bootstrap-t confidence limits can be positioned asymmetric around  $\theta^*$ . Efron and Tibshirani [1993], however, warn that in practice the bootstrap-t can give somewhat erratic results.

An alternative approach to the estimation of a confidence interval suggested by Efron [1982] determines the  $100[\alpha/2]^{\text{th}}$  and  $100[1-\alpha/2]^{\text{th}}$  order values or percentiles directly from the rank-ordered bootstrap estimates  $\theta^*_i$ . An empirical confidence interval could thus be defined by identifying the appropriate [depending on the chosen  $\alpha$ -level] percentiles in the bootstrap sampling distribution without making any parametric assumptions. The empirical confidence interval limits consequently would not necessarily lay symmetrically around  $\theta^*$ . Efron [1982] and Efron and Tibshirani [1986; 1993] refers to this non-parametric approach as the percentile method. However, should the bootstrap estimate  $\theta^*_i$  follow a normal distribution, the percentile and standard methods would deliver equivalent confidence intervals [Efron & Tibshirani, 1986].

A number of modifications to the percentile bootstrap confidence interval discussed above have been suggested [Efron, 1982; Shao & Tu, 1995] to combat the reported tendency [Efron, 1982; Rasmussen, 1987; Strube, 1988] for the aforementioned non-parametric bootstrap methodology to deliver overly restricted confidence intervals and thus overly liberal Type I error rates. Efron [1982] firstly suggests an adjusted bootstrap interval, obtained by extending the percentile bootstrap interval by  $[(n+2)/(n-1)]^{1/2}$  in both directions. As an alternative solution Efron [1982] suggests a bias-corrected confidence

[BC] interval which differs in length from the percentile interval. The latter adjustment procedure modifies the percentile interval for bias in the sample estimate  $\theta^{\wedge}$  by shifting the confidence interval limits depending on the location of the obtained statistic  $[\theta^{\wedge}]$  in relation to the median bootstrap statistic  $[\text{Me}[\theta^{\wedge}_{bi}]]$ . The adjustment is accomplished through the quantity  $z^{\wedge}_0$  representing a bias-correction constant. The bias-correcting constant  $z^{\wedge}_0$  equals the normalised standard score corresponding to the percentile rank of  $\theta^{\wedge}$  in the bootstrap distribution [i.e. the standard score which cuts off an area in the standard normal distribution equal to the area below  $\theta^{\wedge}_{bi}$  in the bootstrap distribution] [Shao & Tu, 1995]. The upper and lower  $[1-\alpha]100\%$  BC confidence interval limits  $[\theta^{\wedge}_{b[lo]}]$  &  $\theta^{\wedge}_{b[up]}$  are given by the bootstrap estimates  $q^{\wedge}_{bi}$  corresponding to  $\alpha_1 100\text{th}$  and  $\alpha_2 100\text{th}$  percentiles in the bootstrap distribution, where  $\alpha_1$  and  $\alpha_2$  are given by Equation 4.7 [Efron & Tibshirani, 1993].

$$\begin{aligned} \alpha_1 &= \Phi\{z^{\wedge}_0 + z_{\alpha/2}\} \quad \Bigg\} \\ \alpha_2 &= \Phi\{z^{\wedge}_0 + z_{1-\alpha/2}\} \quad \Bigg\} \end{aligned} \quad \text{-----} \quad 4.7$$

where  $\Phi$  indicates the standard normal cumulative distribution function; and

$z_{\alpha/2}$  indicates the standard normal score corresponding to a percentile rank of  $100[\alpha/2]$

Efron & Tibshirani [1993], in addition, suggest a bias-corrected and accelerated confidence  $[BC_a]$  interval. The  $BC_a$  interval limits are also given by percentiles of the bootstrap distribution. The upper and lower  $[1-\alpha]100\%$   $BC_a$  confidence interval limits  $[\theta^{\wedge}_{b[lo]}]$  &  $\theta^{\wedge}_{b[up]}$  are given by the bootstrap estimates  $q^{\wedge}_{bi}$  corresponding to  $\alpha_1 100\text{th}$  and  $\alpha_2 100\text{th}$  percentiles in the bootstrap distribution [Shao & Tu, 1995], where  $\alpha_1$  and  $\alpha_2$  are given by Equation 4.8 [Efron & Tibshirani, 1993].

$$\begin{aligned} \alpha_1 &= \Phi\{z^{\wedge}_0 + (z^{\wedge}_0 + z_{\alpha/2}) / (1 - a^{\wedge}[z^{\wedge}_0 + z_{\alpha/2}])\} \quad \Bigg\} \\ \alpha_2 &= \Phi\{z^{\wedge}_0 + (z^{\wedge}_0 + z_{1-\alpha/2}) / (1 - a^{\wedge}[z^{\wedge}_0 + z_{1-\alpha/2}])\} \quad \Bigg\} \end{aligned} \quad \text{-----} \quad 4.8$$

where  $\Phi$  indicates the standard normal cumulative distribution function; and

$z_{\alpha/2}$  indicates the standard normal score corresponding to a percentile rank of  $100[\alpha/2]$

The quantities  $a^{\wedge}$  represents an acceleration constant whereas  $z^{\wedge}_0$  still represent a bias-correction constants. The quantity  $a^{\wedge}$  indicates [Efron & Tibshirani, 1993, p.186]:

... the rate of change of the standard error of  $\hat{\theta}$  with respect to the true parameter value  $\theta$ . The standard normal approximation  $\hat{\theta} \approx N(\theta, se^2)$  assumes that the standard error of  $\hat{\theta}$  is the same for all  $\theta$ . However this is often unrealistic and the acceleration constant  $\hat{a}$  corrects for this.

Although the acceleration constant  $a$  is considered difficult to estimate [Shao & Tu, 1995], Efron & Tibshirani [1993, p. 186] nonetheless offer Equation 4.9 as one possible way of obtaining an approximation based on the jackknife [see ¶4.5.2].

$$\hat{a} = \frac{\{\sum(\theta_j - \hat{\theta}_{ji})^3\}}{\{6[\sum(\theta_j - \hat{\theta}_{ji})]^3/2\}} \quad 4.9$$

where  $\hat{\theta}_{ji}$  represent jackknife estimates of  $\theta$ ; and

$$\hat{\theta}_j = \sum \hat{\theta}_{ji} / n; i=1, 2, \dots, n$$

If  $\hat{a}$  and  $z_0$  equal zero the  $BC_a$  interval becomes the percentile interval since  $\alpha_1$  and  $\alpha_2$  would then equal  $\Phi\{z_{\alpha/2}\}$  and  $\Phi\{z_{1-\alpha/2}\}$  respectively which equals  $\alpha/2$  and  $1-\alpha/2$  [Efron & Tibshirani, 1993]. The effect of all the modifications suggested to the standard percentile procedure, is to reduce the inflation of the Type I error rate above the chosen error rate  $[\alpha]$ .

#### 4.5.1.2 Bootstrap Hypothesis Testing

Investigating the statistical significance of the observed correlation coefficient through the testing of the null hypothesis  $H_0: \rho = 0$  probably constitutes the more prevalent form of hypothesis testing in validation research. Conventional hypothesis testing procedures provide an adequate statistical tool to decide on the validity of  $H_0$ . Bootstrap approaches to hypothesis testing are consequently not essential in the conventional case. By contrast, however, investigations on the statistical significance of corrected correlation coefficients through the testing of  $H_0: \rho_c = 0$  might benefit from the availability of hypothesis testing procedures with the bootstrap. A non-distribution free bootstrap hypothesis test could be conducted by calculating the standard deviation of the distribution of bootstrap estimates, converting the observed  $r_c$  to a standard score and, assuming the test statistic to follow the standard normal distribution [Moran, 1970], determine the conditional probability of the sample result under the assumption of  $H_0$ . A one-sample distribution free bootstrap method to investigate the preceding null hypothesis would require the following steps:

- choose a null distribution for the sample data under  $H_0$ ;
- calculate the appropriate test statistic on the observed corrected correlation;



- generate  $n$  bootstrap samples;
- evaluate the test statistic on each bootstrap sample; and
- estimate the observed exceedence probability  $[\alpha_b]$  or achieved significance level [ASL] and compare to the critical exceedence probability  $[\alpha_c]$ .

Simulating a null distribution for the sample data under  $H_0$ , however, seems to present a practical problem.

Alternatively, a hypothesis test on the significance of a corrected correlation could be performed by constructing a bootstrap confidence interval and by inspecting the interval to see whether it contains zero or not. Should the latter outcome occur, the null hypothesis would be rejected, whereas the first outcome would result in the conclusion that the observed corrected correlation is not significant at the chosen significance level [Efron & Tibshirani, 1993].

#### 4.5.1.3 Bootstrap Sample Size Requirements

The bootstrap is a computer intensive procedure. It in essence substitutes elegant and sophisticated analytical/theoretical argument with brute computational power [Efron, 1982; Thompson 1991]. Efron [1982; 1987; 1988] suggests that as few as ten or twenty, but seldom more than two hundred, bootstrap replications are needed to give an acceptable bootstrap estimate of the standard deviation of the bootstrapped  $\theta^*_b$  distribution [i.e. the standard error of  $\theta^*_b$  as an approximation of  $\sigma[\theta^*]$ ]. Bootstrap confidence intervals, in contrast, require a substantial increase in the number of iterations. In excess of 2000, but seldom less than 1000, iterations are required.

In the case of the corrected correlation coefficient, more than one bootstrap sample per iteration could be required. The precise number of bootstrap samples required per iteration would depend on the measurement design on which the computational version of the correction formula is based. Furthermore, as was indicated earlier, if nested bootstrap sampling is required [to estimate  $\sigma^*[\theta^*_b]$ ] the number of samples drawn will increase dramatically.

#### 4.5.2 Jackknife Estimates Of Statistical Error

Quenouille [in Efron & Gong, 1983] is generally credited with originating the jackknife as a non-parametric device for estimating bias. This initial groundwork was subsequently further developed in



the 1950's by John W Tukey of Princeton University and the Bell Laboratories. The term jackknife was conceived by Tukey [1956] to convey the idea that the method is an all-purpose statistical tool applicable to a wide variety of estimators [Diaconis & Efron, 1983]. Like the bootstrap, the historically older jackknife generates a sampling distribution for the statistic/estimator of interest  $[\theta^{\wedge}]$  by resampling the original sample of size  $n$  [Thompson, 1991]. The jackknife accomplishes this by removing one observation at a time from the original sample and then recalculating the statistic/estimator  $[\theta^{\wedge}_j; j=1, 2, \dots, n]$  for each of the resulting subsamples of size  $[n-1]$  [Diaconis & Efron, 1983; Efron & Gong, 1983; Shao & Tu, 1995]. The variability of the statistic/estimator across all  $[n-1]$  subsamples can then be assessed. The central postulate underlying the jackknife is that the variability of  $\theta^{\wedge}_j$  across the jackknife samples will approximate the variability in  $\theta^{\wedge}$  that would be obtained should  $t \rightarrow \infty$  independent random samples of size  $n$  be selected from the population [i.e.  $\sigma^2[\theta^{\wedge}_j] \cong \sigma^2[\theta^{\wedge}]$ ]. The jackknife estimate of the standard error of  $\theta^{\wedge}$  is defined by Equation 4.10 [Efron & Tibshirani, 1993; Shao & Tu, 1995].

$$\sigma[\theta^{\wedge}] \cong \sigma[\theta^{\wedge}_b] = \{(n-1/n)\Sigma(\theta^{\wedge}_{ji} - (\Sigma\theta^{\wedge}_{ji}/n))^2\}^{1/2} \quad 4.10$$

The jackknife has the advantage over the bootstrap that it requires less calculations than the bootstrap. Generally, however, the bootstrap is considered superior to the jackknife for the estimation of standard errors and confidence intervals [Diaconis & Efron, 1983; Efron, 1982; Thompson, 1991].

## 4.6 EXPRESSIONS FOR THE STANDARD ERROR OF THE PEARSON CORRELATION COEFFICIENT

### 4.6.1 Obtained, Uncorrected First-Order Pearson Correlation Coefficient

The sampling distribution of the Pearson product moment correlation coefficient is dependant on the size of the sample, the population distribution and the value of the population parameter  $\rho$  [Guilford & Fruchter, 1978; Hays, 1973]. When the hypothesis  $H_0: \rho[X,Y] = 0$  is true, and the population can be assumed to be bivariate normal in its distributional form, the sampling distribution of  $\rho^{\wedge}[X,Y]$  approaches a normal  $[0; 1/\sqrt{(n-1)}]$  distribution as sample size increases. However, if sampling from a bivariate normal population with  $\rho[X,Y] = 0$ , the  $t$  ratio shown as Equation 4.11 will be distributed as Student's  $t$  distribution with  $n-2$  degrees of freedom [Guilford & Fruchter, 1978; Hays, 1971].

$$t = (\rho^{\wedge}[X,Y]\sqrt{n-2})/\sqrt{(1-\rho^{\wedge 2}[X,Y])} \quad 4.11$$

If, however, the population correlation  $\rho[X,Y] \neq 0$  the sampling distribution of  $\rho^*[X,Y]$  tends to be skewed, with this tendency intensifying the more  $\rho$  departs from zero. For virtually any value of  $\rho$  in a bivariate normal population, though, Fisher transformations of  $\rho^*[X,Y]$  to  $Z_{\rho^*}$  [shown as Equation 4.12] will be distributed approximately normal with mean  $0,5\ln\{(1+\rho^*[X,Y])/(1-\rho^*[X,Y])\}$  and variance [approximately]  $1/(n-3)$  [Hays, 1971] or, as a slightly less rough approximation,  $\{1/(n-1)\} + \{(4-\rho^2)/(2(n-1)^2)\}$  [Stuart & Kendall, 1977].

$$Z_{\rho^*} = 0,5\ln\{(1+\rho^*[X,Y])/(1-\rho^*[X,Y])\} \text{-----} 4.12$$

The fit of the Gaussian curve improves as the absolute value of  $\rho[X,Y]$  decreases and the sample size increases [Guilford & Fruchter, 1978; Hays, 1971].

Asymptotically [i.e. when the size of samples drawn from  $\Pi$  approach infinity] the variance of the correlation under conditions of a bivariate normal population is shown as Equation 4.13 [Bobko & Rieck, 1980; Kendall & Stuart, 1977].

$$\sigma^2[\rho^*[X,Y]] = ((1-\rho[X,Y])^2)n^{-1} \text{-----} 4.13$$

Stuart and Kendall [1977, p. 251], however, warn:

The use of the standard error to test a hypothetical non-zero value of  $\rho$  is not, however, to be recommended, since the sampling distribution of  $r$  tends to normality very slowly.

When examining the effect of statistical corrections for restriction of range and/or random measurement error applied to the correlation coefficient, the question naturally arises how the sampling behaviour of the correlation is affected by the correction. As the preceding argument indicates, variability represents an important facet of the sampling behaviour of any statistic. However, when comparing the standard error of the attenuated correlation coefficient [i.e. the uncorrected coefficient] to the standard error of the corrected correlation, the [rather crucial] dilemma emerges what estimate of the standard error of the Pearson correlation coefficient should serve as the appropriate benchmark. The choice of the expression in terms of which the standard error of the corrected correlation is estimated will probably significantly affect the conclusions on the effect of the statistical corrections on the sampling behaviour of the corrected coefficient. The square root of the asymptotic variance of the correlation [Equation 4.13] seems to be the expression most often used for comparison purposes within the aforementioned context [Allen & Dunbar, 1990; Bobko, 1983; Bobko & Rieck, 1980; Forsyth & Feldt, 1969; Hakstian, Schroeder & Rogers, 1988; Hakstian, Schroeder & Rogers, 1989]. The square root of Equation 4.13 will consequently also be used in all subsequent analyses when an estimate of the standard error of the uncorrected correlation coefficient is required. The choice of Equation 4.13 as benchmark is, however, not without criticism [Stuart & Kendall, 1977]. The standard

error associated with Equation 4.12 possibly would have been a more satisfactory choice [Schepers, 1996]. The variance of the corrected correlation coefficient transformed to Fisher's Z is, however, not known for most corrections [Mendoza, 1993]. Difficulties thus would arise when trying to compare the two standard errors.

#### 4.6.2 Pearson Correlation Coefficient Corrected For Criterion And Predictor Unreliability

Deriving an exact analytical expression for the sampling variance for the correlation corrected for attenuation  $\rho^*[T_X, T_Y]$  is regarded by Rogers [1976, p. 121] as "exceedingly difficult, if not impossible". Kelley [1947, p. 528], notwithstanding, employing logarithmic differentials, derived an approximate formula for the square of the standard error [i.e. variance error] of the fully disattenuated correlation coefficient shown as Equation 4.14.

$$\sigma^2[\rho^*[T_X, T_Y]] = \rho^2[T_X, T_Y] / [4(n-2)] \{ 4\rho^2[T_X, T_Y] + 4/\rho^2[X, Y] + 1/\rho'^2_{ttX} + 1/\rho'^2_{ttY} - 4/\rho'^2_{ttX} - 4/\rho'^2_{ttY} \} \text{-----} 4.14$$

where  $\rho'_{ttX}$  and  $\rho'_{ttY}$  refer to correlations between essentially parallel halves on the predictor and criterion [i.e. before being stepped up by the Spearman-Brown prophecy formula]

Equation 4.14 assumes that  $\rho[T_X, T_Y]$  is calculated through either Equation 4.15 or Equation 4.16 attributed to Yule [Kelley, 1947, pp. 527-529].

$$\rho[T_X, T_Y] = \rho[X, Y] / \sqrt{(2\rho'_{ttX} / (1 + \rho'_{ttX})) \sqrt{(2\rho'_{ttY} / (1 + \rho'_{ttY}))}} \text{-----} 4.15$$

$$\rho[T_X, T_Y] = (\rho[X_1, Y_1]\rho[X_1, Y_2]\rho[X_2, Y_1]\rho[X_2, Y_2])^{1/4} / (\sqrt{(\rho[X_1, X_2]\rho[Y_1, Y_2])}) \text{-----} 4.16$$

Forsyth and Feldt [1969] empirically investigated the accuracy of the Kelley formula through a series of Monte Carlo simulations and found it to systematically return values marginally smaller than the actual variance in the  $\rho^*[T_X, T_Y]$  distribution. A better approximation of the standard error is obtained if the square root of Equation 4.14 is multiplied with  $\sqrt{[n/(n-1)]}$  [Forsyth & Feldt, 1969; Mayer, 1983]. In addition Forsyth and Feldt [1969] report that:

- the standard error associated with the corrected correlation coefficient is larger than the standard error of the uncorrected coefficient;
- the skewness of the sampling distribution is negative when the parameter is relatively low;

- the skewness of the sampling distribution tends to increase as the parameter increases up to 0,80;
- at  $\rho[T_X, T_Y]=1$  the trend of progressive skewness reverses and the sampling distribution tends to be slightly positively skewed;
- the sampling distribution tends to be leptokurtic for  $\rho[T_X, T_Y]>0,30$ ;
- the sampling distribution approximates the Gaussian distribution for "suitably large samples".

Forsyth and Feldt [1969] indicate that the use of the Kelly [1947] approximation of  $\sqrt{\sigma^2[\rho^*[T_X, T_Y]]}$  in a routine Gaussian curve procedure for establishing confidence intervals works quite well, provided that  $n$  at least exceeds 100. A procedure that provides satisfactory control of Type I error is thus also indicated for testing the significance of correlations corrected for the attenuating effect of predictor and criterion unreliability.

Mayer [1983] presents a formula [shown as Equation 4.17] for estimating the sampling variance of the correlation fully corrected for attenuation when Cronbach's alpha was used to calculate the reliability estimates for the predictor and the criterion.

$$\sigma^2[\rho^*[T_X, T_Y]] = \rho^2[T_X, T_Y]/[(n-2)]\{\rho^2[T_X, T_Y]+1/\rho^2[X, Y]+1/\rho^2_{ttX}+1/\rho^2_{ttY}-3/\rho_{ttX}^3-3/\rho_{ttY}^3+2\} \text{-----} 4.17$$

Equation 4.16 represents a modification of the Kelley [1947] expression to accept coefficient alpha estimates directly [Mayer, 1983].

Hakstian, Schroeder and Rogers, [1988, p. 31] employed the delta method to develop an asymptotic expression for the square of the standard error of the fully corrected correlation coefficient shown as Equation 4.18.

$$\sigma^2[\rho^*[T_X, T_Y]] = A + B + C - D + E - F + G + H \text{-----} 4.18$$

$$A = \sigma^2[\rho^*[X_1, Y_2]] + \sigma^2[\rho^*[X_2, Y_1]] + 2\{\sigma[\rho^*[X_1, Y_2], \rho^*[X_2, Y_1]]\}/(4\rho[X_1, X_2]\rho[Y_1, Y_2])$$

$$B = \{(\rho[X_1, Y_2] + \rho[X_2, Y_1])^2/16(\rho[X_1, X_2])^3\rho[Y_1, Y_2]\}\{\sigma^2[\rho^*[X_1, X_2]]\}$$

$$C = \{(\rho[X_1, Y_2] + \rho[X_2, Y_1])^2/16(\rho[X_1, X_2])\rho[Y_1, Y_2]^3\}\{\sigma^2(\rho^*[Y_1, Y_2])\}$$

$$D = \{(\rho[X_1, Y_2] + \rho[X_2, Y_1])/4\rho[X_1, X_2]^2\rho[Y_1, Y_2]\}\{\sigma(\rho^*[X_1, Y_2], \rho^*[X_1, X_2])\}$$

$$E = \sigma[\rho^*[X_2, Y_1], \rho^*[X_1, X_2]]$$

$$F = \{(\rho[X_1, Y_2] + \rho[X_2, Y_1])/4\rho[X_1, X_2]\rho[Y_1, Y_2]\} \{\sigma(\rho^*[X_1, Y_2], \rho^*[Y_1, Y_2])\}$$

$$G = \sigma[\rho^*[X_2, Y_1], \rho^*[Y_1, Y_2]]$$

$$H = \{(\rho[X_1, Y_2] + \rho[X_2, Y_1])^2\} / \{8(\rho[X_1, X_2]\rho[Y_1, Y_2])^2\} \{\sigma(\rho^*[X_1, X_2], \rho^*[Y_1, Y_2])\}$$

Equation 4.18 assumes that  $\rho_{tt}X$  and  $\rho_{tt}Y$  represent test-retest reliability coefficients. The expression derived by Hakstian, Schroeder and Rogers, [1988] is derived similarly to the formula [Equation 4.19] developed by Kristof [1982]. Kristof [1982, p. 109] derives the following expression for the asymptotic variance of  $\rho^*[T_X, T_Y]\sqrt{n}$ :

$$\sigma^2[\rho^*[T_X, T_Y]\sqrt{n}] = \rho^2[T_X, T_Y] \left\{ \rho^2[T_X, T_Y] + (t_{22}/(t_{11}-t_{22}))^2 + (t_{44}/(t_{33}-t_{44}))^2 - (t_{11}/(t_{11}-t_{22})) - (t_{33}/(t_{33}-t_{44})) + ((t_{11}t_{33})/t^2_{13}) \right\} \text{-----} 4.19$$

Where:

$$t_{11} = s_{11} + s_{22} + 2s_{12}$$

$$t_{33} = s_{33} + s_{44} + 2s_{34}$$

$$t_{13} = s_{13} + s_{14} + s_{23} + s_{24}$$

$$t_{22} = s_{11} + s_{22} - 2s_{12}$$

$$t_{44} = s_{33} + s_{44} - 2s_{34}$$

and  $\sigma_{ij}$  represent covariance terms in a 4x4 matrix defined by two essentially parallel parts  $[Y_1 \text{ \& } Y_2]$  of Y and two essentially parallel parts  $[X_1 \text{ \& } X_2]$  of X

The expressions are not equivalent, however, due to differences in the measurement design in terms of which the correlation coefficients corrected for predictor and criterion unreliability are estimated. The Hakstian, Schroeder and Rogers, [1988] expression presupposes a design in which two measures X and Y are administered at time 1 to yield  $X_1$  and  $Y_1$  but then re-administered after a suitable period of time to yield  $X_2$  and  $Y_2$ . In terms of this design an estimate of the fully disattenuated correlation coefficient  $\rho[T_X, T_Y]$  is obtained through Equation 4.20.

$$\rho[T_X, T_Y] = (\rho[X_1, Y_2] + \rho[X_2, Y_1]) / \{2(\rho[X_1, X_2]^{1/2} \rho[Y_1, Y_2]^{1/2})\} \text{-----} 4.20$$

The Kristof [1982] formula, in contrast, assumes a measurement design in which two measures X and Y have been applied once and subsequently divided into two essentially equivalent parts [X<sub>1</sub> & X<sub>2</sub>, and Y<sub>1</sub> & Y<sub>2</sub>]. Kristof [1982] also examines the asymptotic variance of  $\rho^*[T_X, T_Y]$  if the less stringent assumption is made that the test and criterion had been divided into two congeneric instead of essentially parallel parts

#### 4.6.3 Pearson Correlation Coefficient Corrected For Criterion Unreliability Only

Kelley [1947, p. 529], assuming the computational formula to be Equation 4.21, derived, through logarithmic differentials, an approximate formula for the square of the standard error [i.e. variance error] of the partially disattenuated correlation coefficient shown as Equation 4.22.

$$\rho[X, T_Y] = \rho[X, Y] / \sqrt{2\rho'_{ttY} / (1 + \rho'_{ttY})} \text{-----} 4.21$$

$$\sigma^2[\rho^*[X, T_Y]] = \rho^2[X, T_Y] / (n-2) (\rho^2[X, T_Y] + \rho^{-1}[X, Y] \rho'^{-1}_{ttY} + 4\rho'^{-2}_{ttY} / 4) \text{-----} 4.22$$

where  $\rho'_{ttX}$  and  $\rho'_{ttY}$  refer to correlations between essentially parallel halves on the predictor and criterion [i.e. before being stepped up by the Spearman-Brown prophecy formula]

The reaction of Kelley's approximation of the standard error of the partially disattenuated correlation coefficient to changes in the attenuated correlation, the reliability coefficient and sample size is graphically depicted in Figure 4.1 and Figure 4.2. Figures 4.1 -4.2 reveal that the standard error:

- decreases curvilinearly as  $\rho_{ttY}$  increases;
- increases as the attenuated correlation increases at low values of  $\rho_{ttY}$  but decreases as the attenuated correlation increases at low values of  $\rho_{ttY}$ ;
- decreases as n increases for any  $\rho[X, Y]$   $\rho_{ttY}$  combination.

Figure 4.3 and Figure 4.4 depict the reaction of the standard error of the partially disattenuated correlation expressed in terms of the standard error of the uncorrected coefficient [i.e.  $V = \sigma[\rho^*] / \sigma[\rho^*]$ ] to changes in  $\rho[X, Y]$ ,  $\rho_{ttY}$  and n. Figures 4.3 - 4.4 suggest V to increase as the attenuated correlation increases, but with the rate of increase decreasing as  $\rho_{ttY}$  increases. V also seems to be negatively related to sample size.



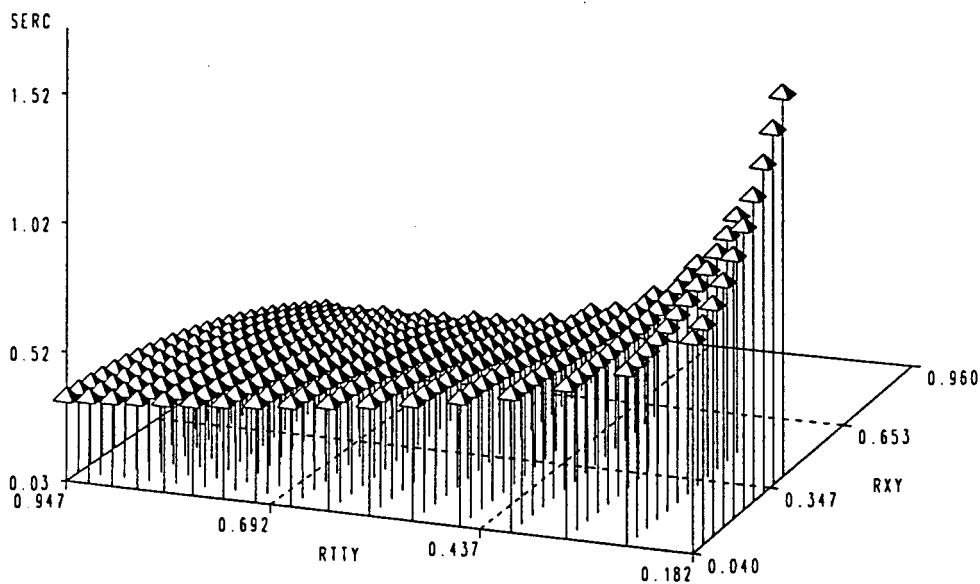


Figure 4.1: The standard error of the partially disattenuated validity coefficient as a function of  $\rho[X,Y]$  and  $\rho_{tt}Y$  for  $n$  fixed at 10

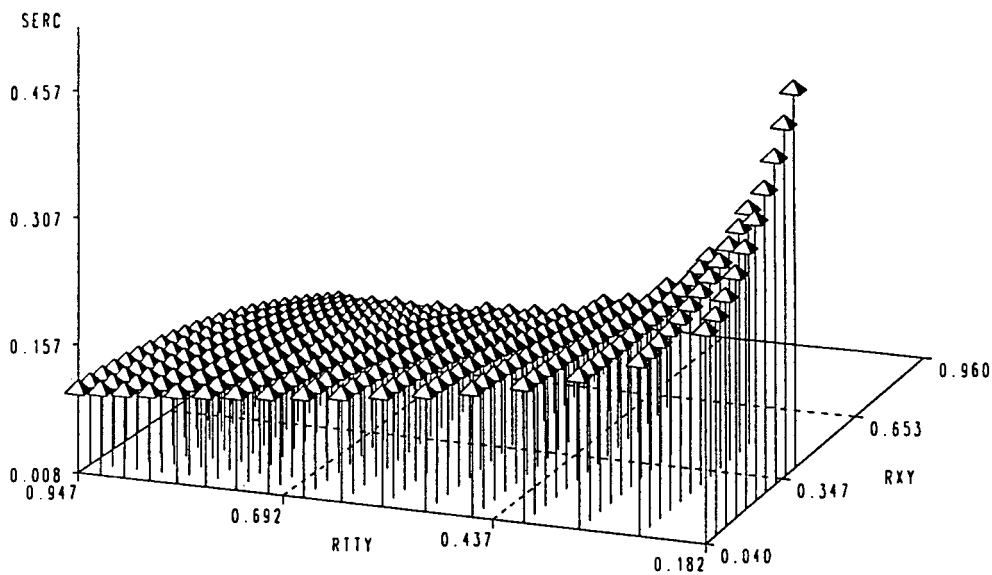


Figure 4.2: The standard error of the partially disattenuated validity coefficient as a function of  $\rho[X,Y]$  and  $\rho_{tt}Y$  for  $n$  fixed at 90

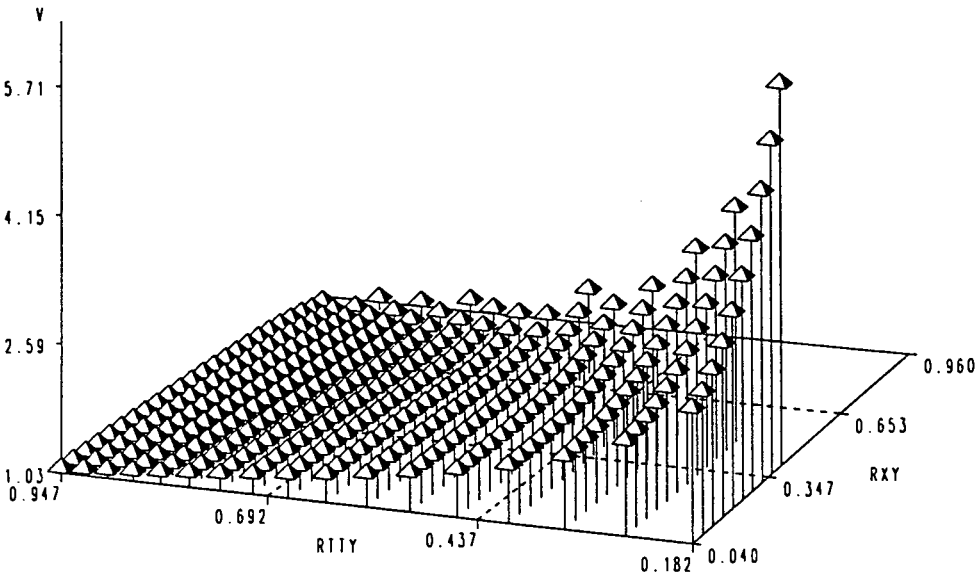


Figure 4.3:  $V = \sigma[\hat{\rho}_c]/\sigma[\hat{\rho}]$  as a function of  $\rho[X,Y]$  and  $\rho_{tt}Y$  for  $n$  fixed at 10

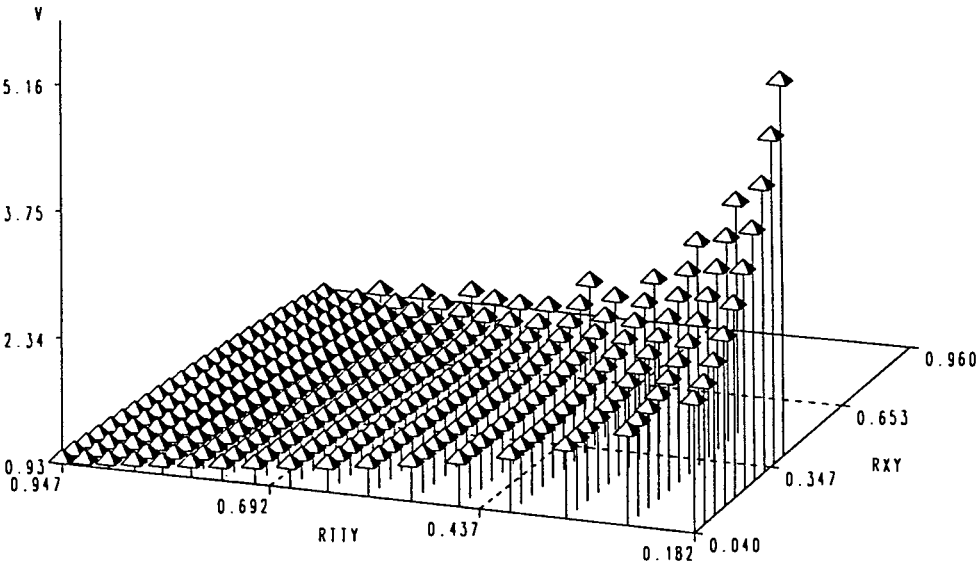


Figure 4.4:  $V = \sigma[\hat{\rho}_c]/\sigma[\hat{\rho}]$  as a function of  $\rho[X,Y]$  and  $\rho_{tt}Y$  for  $n$  fixed at 90

An interesting habit of  $V$  which Figure 4.4 manages to uncover is that  $V$  drops below 1 [i.e. the standard error of the corrected coefficient is in fact smaller than that of the attenuated correlation] in a small region characterised by high  $\rho[X,Y]$  and  $\rho_{ttY}$  values. As  $n$  increases the region gradually seems to expand to lower  $\rho[X,Y]$  and  $\rho_{ttY}$  values. Whether this habit is attributable to Equation 4.13's known weakness in the upper region of the  $\rho[X,Y]$  scale, is uncertain.

Bobko & Rieck [1980, p. 388] derived, through the delta method, expressions for an approximation of the square of the standard error of a partially disattenuated correlation coefficient shown as Equations 4.23, 4.24 and 4.25. Three different formula are presented as a function of the measurement design, specifically the method used to calculate the criterion reliability coefficient;

- $\rho_{ttY}$  obtained from the same data set as  $\rho[X,Y]$  [Case I; Equation 4.23];
- $\rho_{ttY}$  obtained from an independent data set [Case II; Equation 4.24]; and
- $\rho_{ttY}$  accepted a priori on theoretical grounds or previously accepted knowledge [Case III; Equation 4.25].

$$\sigma^2[\rho^*[X,T_Y]] = [A+B-C] \text{-----} 4.23$$

$$\sigma^2[\rho^*[X,T_Y]] = [A+B] \text{-----} 4.24$$

$$\sigma^2[\rho^*[X,T_Y]] = [A] \text{-----} 4.25$$

Where:

$$A = (n)^{-1}(\rho_{ttY})^{-1}(1-\rho^2[X,Y])^2;$$

$$B = (0,25n)^{-1}(\rho_{ttY})^{-3}(\rho[X,Y])^2(1-\rho_{ttY})^2; \text{ and}$$

$$C = (n)^{-1}(\rho_{ttY})^{-2}\rho[X,Y]\{\rho_{ttY}(1-2\rho^2[X,Y])-(0,5\rho^2[X,Y])(1-2\rho^2[X,Y]-\rho_{ttY}^2)\}$$

Figures 4.5 - 4.10 depict the reaction of the standard error of the partially disattenuated correlation coefficient calculated through Equations 4.23 - 4.25 to changes in the attenuated correlation, the reliability coefficient and sample size. When  $\rho_{ttY}$  is obtained from the same group as  $\rho[X,Y]$ , Figures 4.5 - 4.6 suggest that  $\sigma[\rho^*_c]$ :

- curvilinearly increases as the attenuated correlation increases at low values of  $\rho_{ttY}$  but gradually reverses the trend as  $\rho_{ttY}$  increases until at high values of  $\rho_{ttY}$ ,  $\sigma[\rho^*_c]$  moderately decreases as  $\rho[X,Y]$  increases;
- increases as  $\rho_{ttY}$  decreases;
- peaks at the maximum  $\rho[X,Y]$  value possible, given  $\rho_{ttY}$ ;

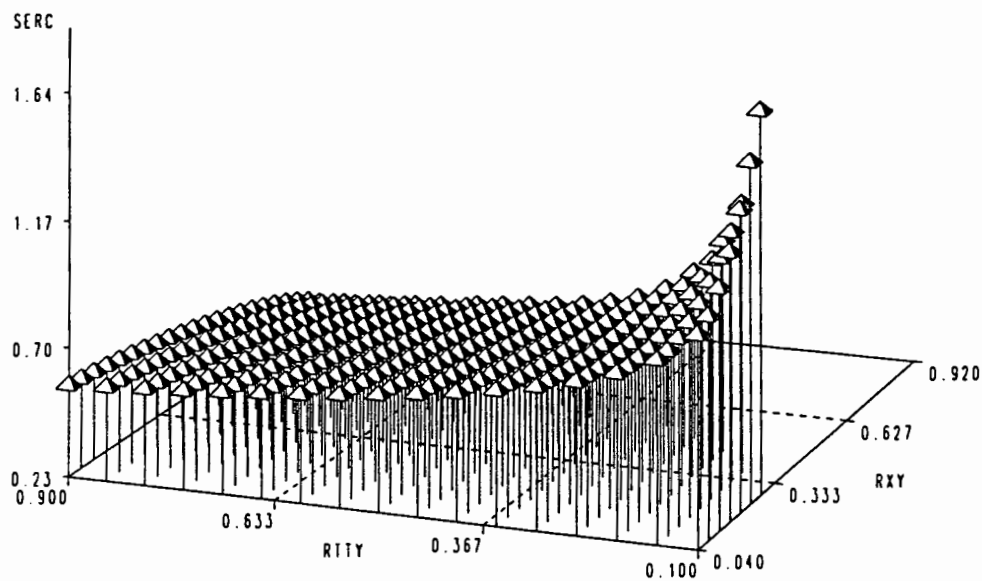


Figure 4.5: The standard error of the partially disattenuated validity coefficient calculated through Equation 4.23 as a function of  $\rho[X,Y]$  and  $\rho_{tt}Y$  for  $n$  fixed at 10

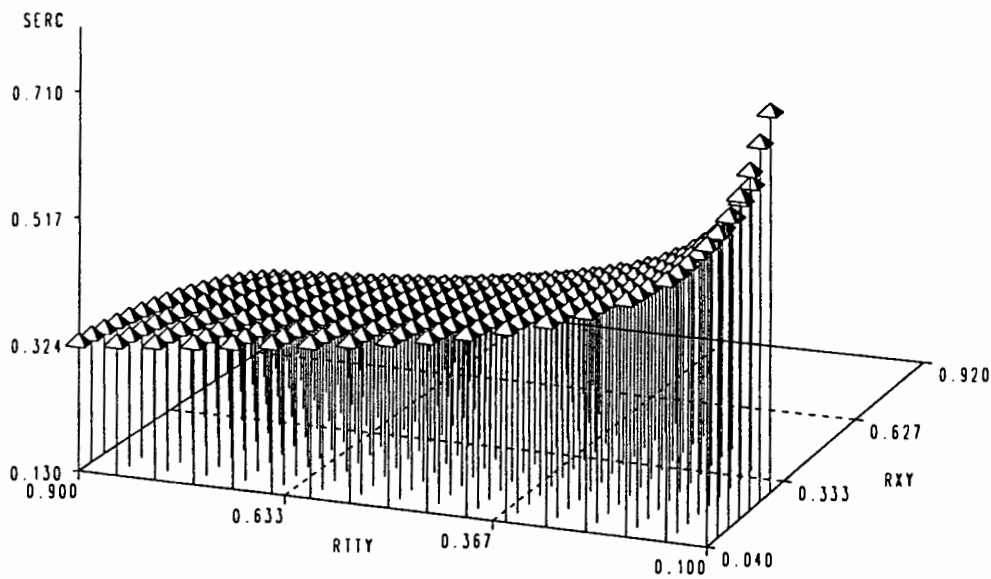


Figure 4.6: The standard error of the partially disattenuated validity coefficient calculated through Equation 4.23 as a function of  $\rho[X,Y]$  and  $\rho_{tt}Y$  for  $n$  fixed at 90

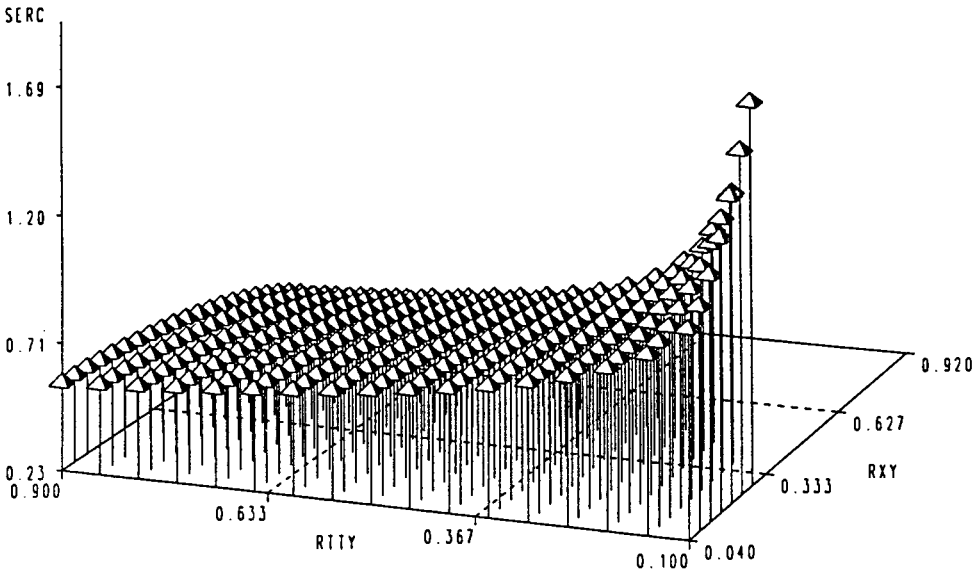


Figure 4.7: The standard error of the partially disattenuated validity coefficient calculated through Equation 4.24 as a function of  $\rho[X,Y]$  and  $\rho_{tt}Y$  for  $n$  fixed at 10

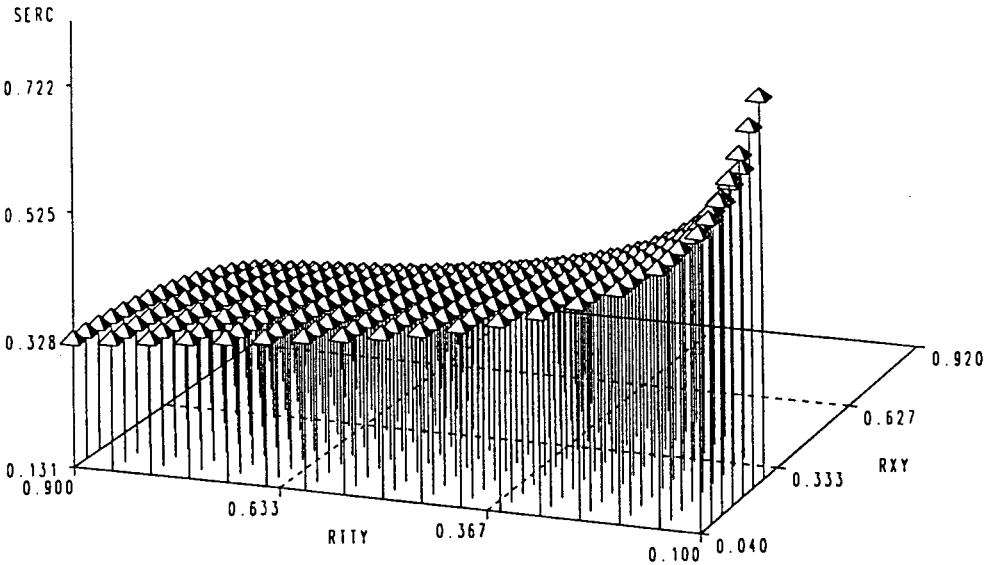


Figure 4.8: The standard error of the partially disattenuated validity coefficient calculated through Equation 4.24 as a function of  $\rho[X,Y]$  and  $\rho_{tt}Y$  for  $n$  fixed at 90

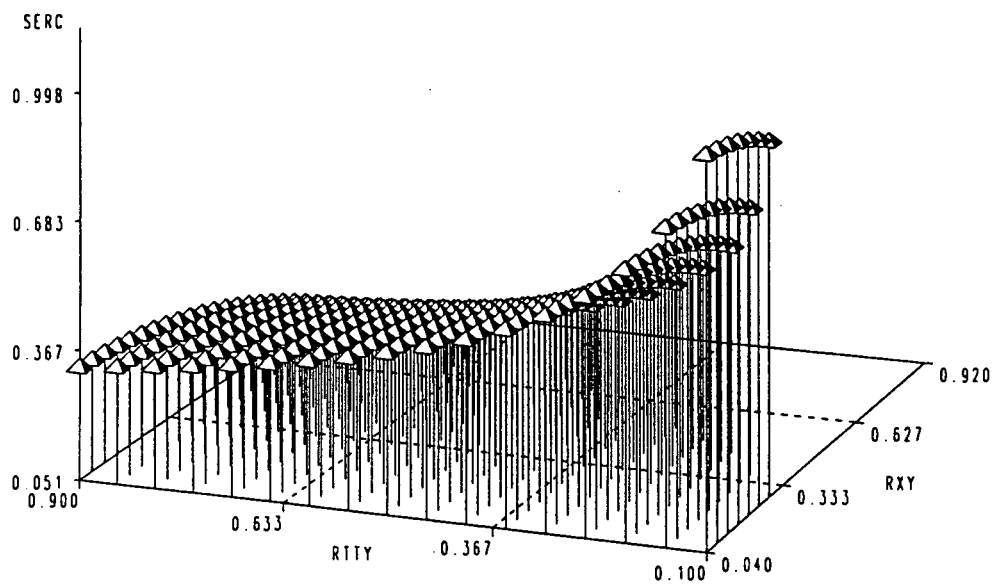


Figure 4.9: The standard error of the partially disattenuated validity coefficient calculated through Equation 4.25 as a function of  $\rho[X,Y]$  and  $\rho_{tt}Y$  for  $n$  fixed at 10

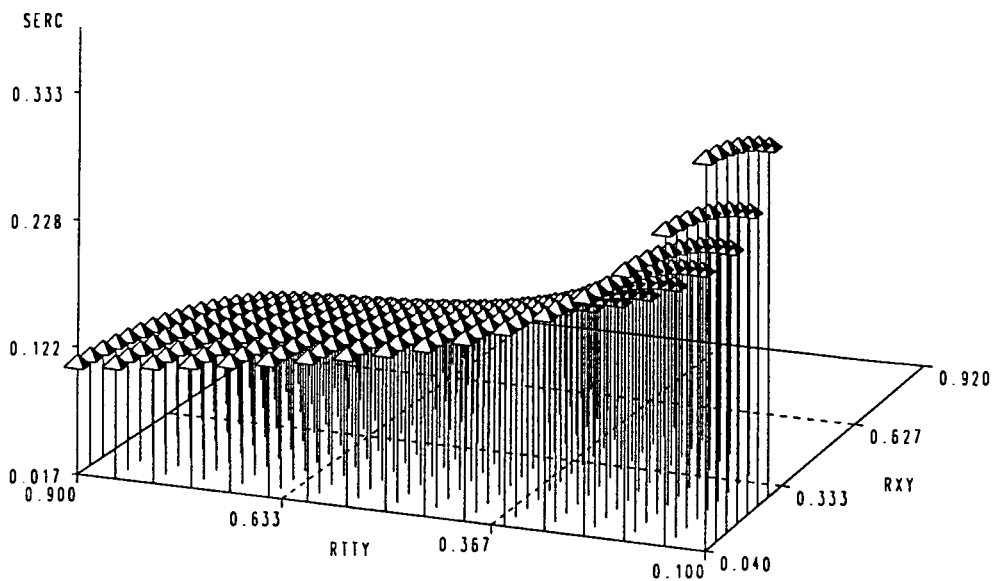


Figure 4.10: The standard error of the partially disattenuated validity coefficient calculated through Equation 4.25 as a function of  $\rho[X,Y]$  and  $\rho_{tt}Y$  for  $n$  fixed at 90



- decreases as sample size increases.

The same conclusions also seem to apply to Figures 4.7 - 4.8 portraying the behaviour of  $\sigma[\rho^{\wedge}_c]$  when  $\rho_{ttY}$  is obtained from an independent group. In contrast, the behaviour of  $\sigma[\rho^{\wedge}_c]$  when  $\rho_{ttY}$  is accepted a priori on theoretical grounds or previously accepted knowledge, seems to be slightly dissimilar to that observed under the two previous measurement designs. Figures 4.9 - 4.10 indicate that  $\sigma[\rho^{\wedge}_c]$  still increases as  $\rho_{ttY}$  decreases and that  $\sigma[\rho^{\wedge}_c]$  still decreases as  $n$  increases. However,  $\sigma[\rho^{\wedge}_c]$  reacts differently to an increase in the attenuated correlation. Although still curvilinearly related to  $\rho[X,Y]$ ,  $\sigma[\rho^{\wedge}_c]$  now seems to gradually decrease as  $\rho[X,Y]$  increases for all values of  $\rho_{ttY}$ .

The behaviour of the ratio of the two standard errors,  $V$  is portrayed in Figures 4.11 - 4.16. For all three measurement designs  $V > 1$ , which implies that  $\sigma[\rho^{\wedge}_c]$  exceeds  $\sigma[\rho^{\wedge}]$  for all [permissible] combinations of values of  $\rho[X,Y]$ ,  $\rho_{ttY}$  and  $n$ . For the first two measurement designs [Case I & II] the reaction of  $V$  to changes in  $\rho[X,Y]$ ,  $\rho_{ttY}$  and  $n$  is again [as would logically be expected] very similar. Figures 4.11 and 4.12 [Case I] and Figures 4.13 and 4.14 [Case II] indicate that the difference between  $\sigma[\rho^{\wedge}_c]$  and  $\sigma[\rho^{\wedge}]$  increases as  $\rho[X,Y]$  increases and as  $\rho_{ttY}$  decreases.  $V$  seems to peak both at high  $\rho_{ttY}$  and high  $\rho[X,Y]$  and at low  $\rho_{ttY}$  and low-moderate  $\rho[X,Y]$ . Figures 4.15 - 4.16, however, show that for the third measurement design [Case III],  $V$  remains constant over changes in  $\rho[X,Y]$  for all values of  $\rho_{ttY}$  and increases as  $\rho_{ttY}$  decreases. This would imply that  $\sigma[\rho^{\wedge}]$  shadows/tails  $\sigma[\rho^{\wedge}_c]$  over all values of  $\rho_{ttY}$  and  $\rho[X,Y]$  but that the latter tends to gain on the former as criterion reliability decreases.

Kristof [1982, p. 101] used the delta method to obtain an asymptotic expression of the sampling variance of the partially disattenuated correlation coefficient [shown as Equation 4.26] when the criterion can be divided into two essentially parallel parts.

$$\sigma^2[\rho^{\wedge}[X,T_Y]\sqrt{n}] = \rho^2[X,T_Y](\rho^2[X,T_Y]-1+(t_{22}/t_{11}-t_{22})^2-(t_{11}/t_{11}-t_{22})+(t_{11}t_{33}/t^2_{13})) \text{-----} 4.26$$

Where:

$t_{ij}$  represent covariance terms in a 3x3 covariance matrix defined by  $Z_1=Y_1+Y_2$ ,  $Z_2=Y_1-Y_2$  &  $Z_3=X$

$$t_{11} = \sigma_{11} + \sigma_{22} + 2\sigma_{12}$$

$$t_{22} = \sigma_{11} + \sigma_{22} - 2\sigma_{12}$$

$$t_{33} = \sigma_{33}$$

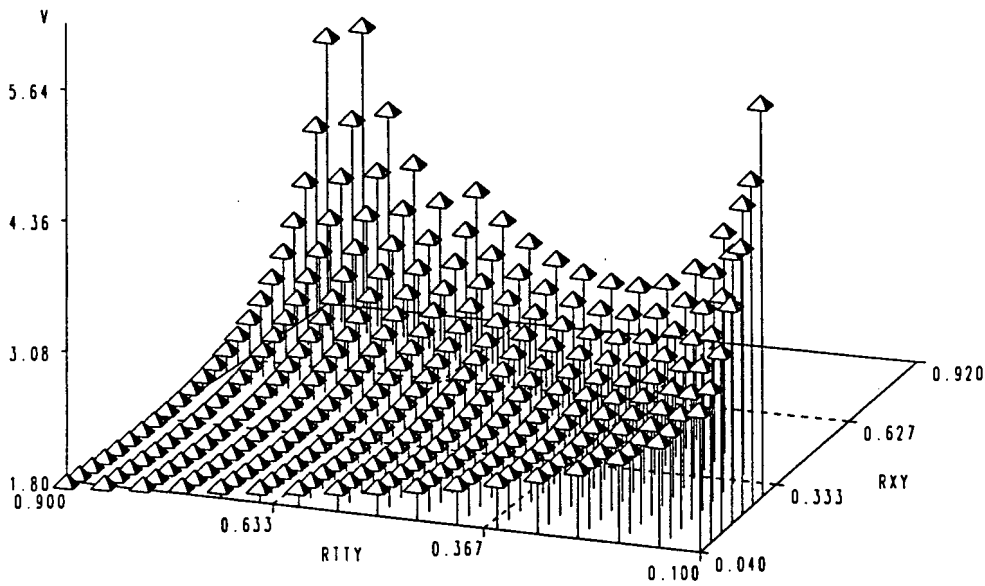


Figure 4.11:  $V = \sigma[\hat{\rho}_c]/\sigma[\hat{\rho}]$  as a function of  $\rho[X,Y]$ ,  $\rho_{tt}Y$  and  $n$  when the standard error of the partially disattenuated validity coefficient is calculated through Equation 4.23 for  $n$  fixed at 10

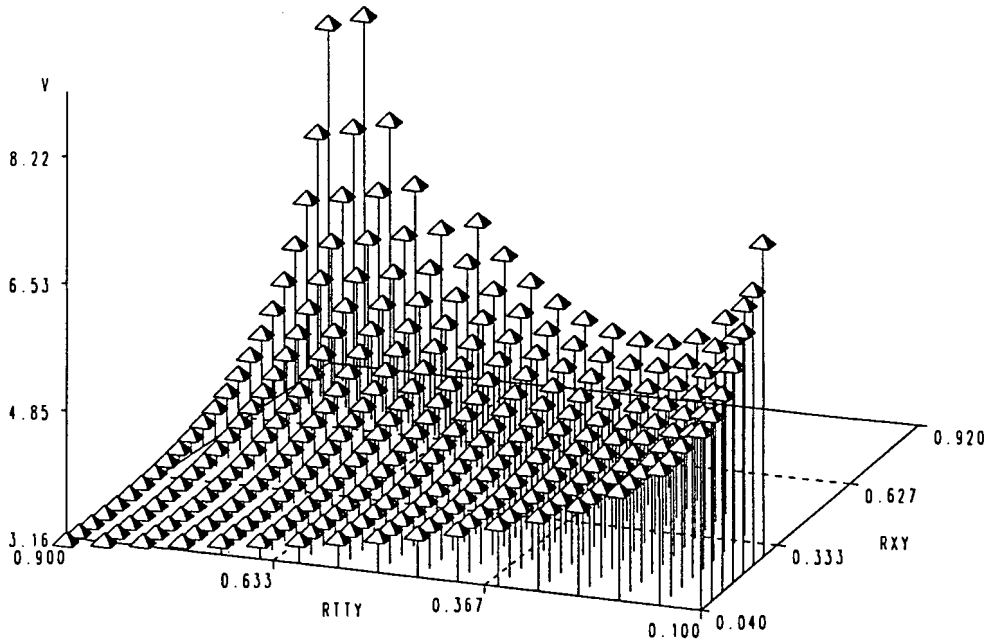


Figure 4.12:  $V = \sigma[\hat{\rho}_c]/\sigma[\hat{\rho}]$  as a function of  $\rho[X,Y]$ ,  $\rho_{tt}Y$  and  $n$  when the standard error of the partially disattenuated validity coefficient is calculated through Equation 4.23 for  $n$  fixed at 90

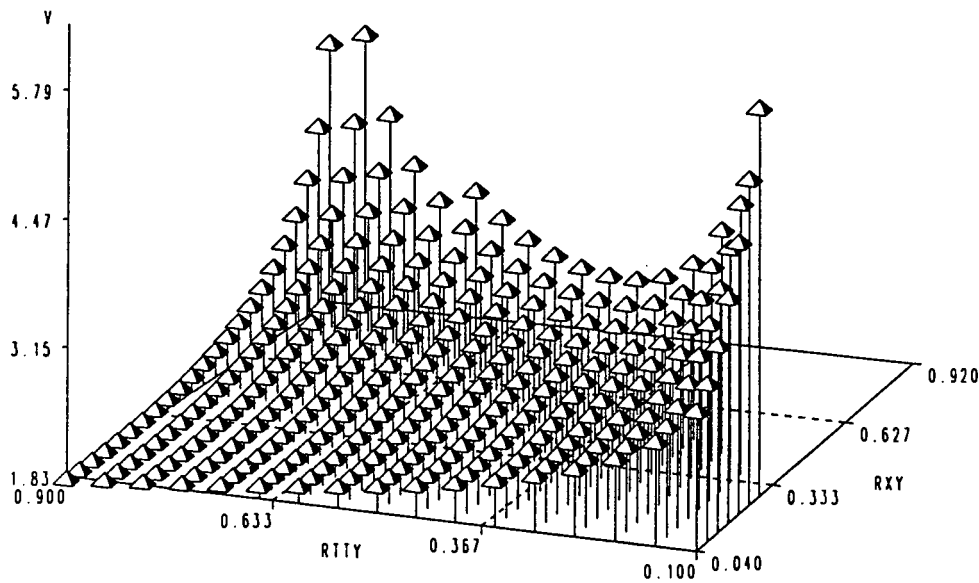


Figure 4.13:  $V = \sigma[\hat{\rho}_c]/\sigma[\hat{\rho}]$  as a function of  $\rho[X,Y]$ ,  $\rho_{tt}Y$  and  $n$  when the standard error of the partially disattenuated validity coefficient is calculated through Equation 4.24 for  $n$  fixed at 10

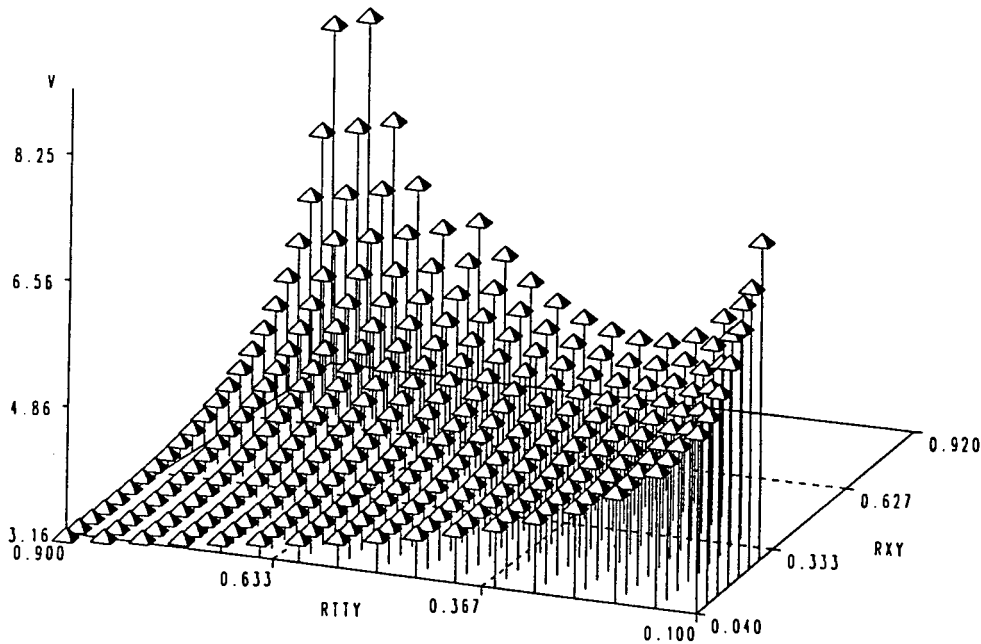


Figure 4.14:  $V = \sigma[\hat{\rho}_c]/\sigma[\hat{\rho}]$  as a function of  $\rho[X,Y]$ ,  $\rho_{tt}Y$  and  $n$  when the standard error of the partially disattenuated validity coefficient is calculated through Equation 4.24 for  $n$  fixed at 90

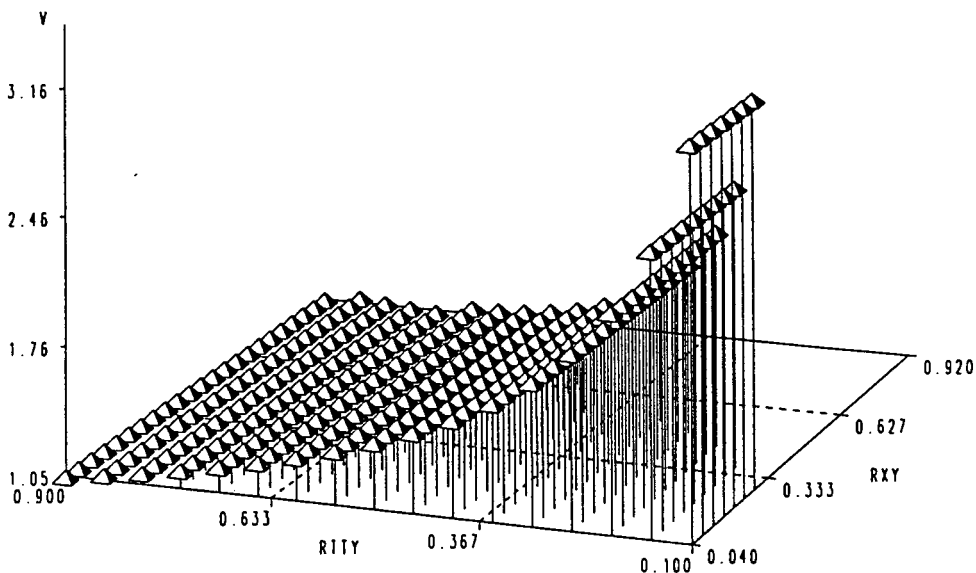


Figure 4.15:  $V = \sigma[\hat{\rho}_{\epsilon}]/\sigma[\hat{\rho}]$  as a function of  $\rho[X,Y]$ ,  $\rho_{ttY}$  and  $n$  when the standard error of the partially disattenuated validity coefficient is calculated through Equation 4.25 for  $n$  fixed at 10

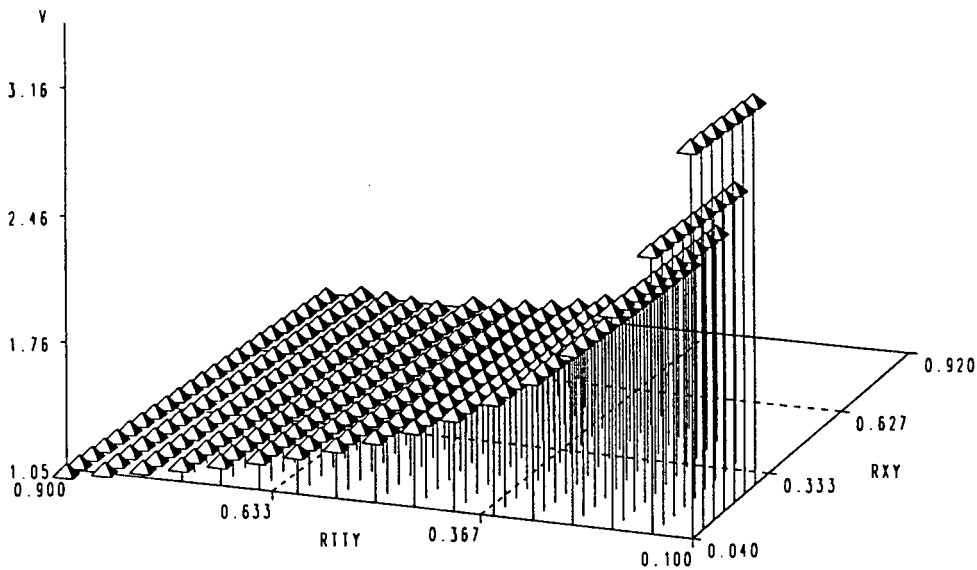


Figure 4.16:  $V = \sigma[\hat{\rho}_{\epsilon}]/\sigma[\hat{\rho}]$  as a function of  $\rho[X,Y]$ ,  $\rho_{ttY}$  and  $n$  when the standard error of the partially disattenuated validity coefficient is calculated through Equation 4.25 for  $n$  fixed at 90

$$t_{13} = \sigma_{13} + \sigma_{23}$$

and:

$\sigma_{ij}$  represent covariance terms in a 3x3 matrix defined by two essentially parallel parts  $[Y_1 \text{ \& } Y_2]$  of Y and X

Hakstian, Schroeder and Rogers [1989, p. 397] regard the Bobko and Rieck [1980] standard error expressions as "insufficiently detailed for general use", probably due to their failure to fully explicate the measurement designs underlying their derivations. Hakstian, Schroeder and Rogers [1989] derive asymptotic expressions for the partially disattenuated correlation coefficient for the following four, more precisely detailed, measurement designs in terms of which  $\rho[X, T_Y]$  is estimated:

- Case I design: one predictor [X] and two criterion  $[Y_1 \text{ \& } Y_2]$  measurements concurrently obtained from a single data set with  $\rho[X, Y]$  being calculated as the mean of  $\rho[X, Y_1]$  and  $\rho[X, Y_2]$  or as  $\rho[X, (Y_1 + Y_2)]$  and  $\rho_{tt}Y$  being estimated by  $\rho[Y_1, Y_2]$ ;
- Case II design: one predictor [X] and two criterion  $[Y_1 \text{ \& } Y_2]$  measurements obtained [with the same time interval between the X and  $Y_2$ , and  $Y_1$  and  $Y_2$  measurements] from a single data set with  $\rho[X, Y]$  estimated by  $\rho[X, Y_2]$  and  $\rho_{tt}Y$  estimated by  $\rho[Y_1, Y_2]$ ;
- Case III design: one predictor [X] and one criterion [Y] measurement obtained from data set A and two criterion measurements  $[Y_1 \text{ \& } Y_2]$  obtained from a second, independent data set B [of comparable heterogeneity] [with the same time interval between the X and Y, and  $Y_1$  and  $Y_2$  measurements] with an estimate of  $\rho[X, Y]$  calculated from sample A and an estimate of  $\rho_{tt}Y$  calculated from sample B; and
- Case IV design: one predictor [X] and one criterion [Y] measurement obtained from data set A with an estimate of  $\rho[X, Y]$  calculated from sample A and  $\rho_{tt}Y$  taken as known from previous theory or from a test manual, with the proviso that the between measurements time interval lengths and group heterogeneity of the observed sample and the standardisation sample match.

The four measurement designs imply the following four computational formula shown as Equations 4.27, 4.28, 4.29 and 4.30 [Hakstian, Schroeder & Rogers, 1989]:

For a Case I design:

$$\rho[X, T_Y] = (\rho[X, Y_1] + \rho[X, Y_2]) / 2(\rho[Y_1, Y_2]^{1/2}) \text{-----} 4.27$$

For a Case II design:

$$\rho[X, T_Y] = \rho[X, Y_2] / (\rho[Y_1, Y_2])^{1/2} \text{-----} 4.28$$

For a Case III design:

$$\rho[X, T_Y] = \rho[X, Y] / (\rho[Y_1, Y_2])^{1/2} \text{-----} 4.29$$

For a Case IV design:

$$\rho[X, T_Y] = \rho[X, Y] / (\rho[Y_1, Y_2])^{1/2} \text{-----} 4.30$$

Asymptotic expressions for the square of the standard error of the partially disattenuated correlation coefficient estimated by Hakstian, Schroeder and Rogers [1989, pp. 399-401] via the foregoing four measurement designs/computational formula are shown as Equations 4.31, 4.32, 4.33 and 4.34.

For a Case I design:

$$\begin{aligned} \sigma^2[\rho^{\wedge}[X, T_Y]] = & \{(\sigma^2[\rho^{\wedge}[X, Y_1]] + \sigma^2[\rho^{\wedge}[X, Y_2]]) / (4\rho[Y_1, Y_2])\} + \\ & \{((\rho[X, Y_1] + \rho[X, Y_2])^2 \sigma^2[\rho^{\wedge}[Y_1, Y_2]]) / 16\rho^3[Y_1, Y_2]\} + \\ & \{\sigma[\rho^{\wedge}[X, Y_1], \rho^{\wedge}[X, Y_2]] / 2\rho[Y_1, Y_2]\} - \\ & \{(\rho[X, Y_1] + \rho[X, Y_2])(\sigma[\rho^{\wedge}[X, Y_1], \rho^{\wedge}[Y_1, Y_2]] + \\ & \sigma[\rho^{\wedge}[X, Y_2], \rho^{\wedge}[Y_1, Y_2]]) / 4\rho^2[Y_1, Y_2]\} \text{-----} 4.31 \end{aligned}$$

For a Case II design:

$$\begin{aligned} \sigma^2[\rho^{\wedge}[X, T_Y]] = & \{\sigma^2[\rho^{\wedge}[X, Y_2]] / \rho[Y_1, Y_2]\} + \\ & \{\rho^2[X, Y_2] \sigma^2[\rho^{\wedge}[Y_1, Y_2]] / 4\rho^3[Y_1, Y_2]\} - \\ & \{\rho[X, Y_2] \sigma[\rho^{\wedge}[X, Y_2], \rho[Y_1, Y_2]] / \rho^2[Y_1, Y_2]\} \text{-----} 4.32 \end{aligned}$$

For a Case III design:

$$\begin{aligned} \sigma^2[\rho^{\wedge}[X, T_Y]] = & \{\sigma^2[\rho^{\wedge}[X, Y]^{\text{A}}] / \rho[Y_1, Y_2]^{\text{B}}\} + \\ & \{(\rho^2[X, Y]) \sigma^2[\rho^{\wedge}[Y_1, Y_2]^{\text{B}}] / 4(\rho^3[Y_1, Y_2]^{\text{B}})\} \text{-----} 4.33 \end{aligned}$$

For a Case IV design:

$$\sigma^2[\rho^{\wedge}[X, T_Y]] = \sigma^2[\rho^{\wedge}[X, Y]] / \rho_{ttY} \text{-----} 4.34$$



#### 4.6.4 Pearson Correlation Coefficient Corrected For Case 1 [Case B] Restriction Of Range

No formal derivation of a specific expression for the standard error of the correlation coefficient corrected for Case 1 [Case B] selection could be uncovered. Validation researchers thus apparently have no option but to rely on resampling techniques [e.g. bootstrapping] to generalise their corrected correlational sample results beyond the confines of the sample studied.

#### 4.6.5 Pearson Correlation Coefficient Corrected For Case 2 [Case A] Restriction Of Range

Bobko and Rieck [1980] used the delta method to derive an expression for an approximation of the square of the standard error of a correlation coefficient corrected for explicit restriction of range on the predictor, shown as Equations 4.35.

$$\sigma^2[\rho^{\wedge}_c[X,Y]] = \{K^2(1-\rho^{\wedge 2}[x,y]+\rho^{\wedge 2}[x,y]K^2)^{-3}\} \{(1-\rho^{\wedge 2}[x,y])^2/n\} \text{-----} 4.35$$

Where:

$$K = (\sigma[X]/\sigma[x])$$

Bobko and Rieck [1980] point out that Equation 4.35 is identical to the expression for the square of the standard error of the correlation corrected for Case 2 [Case A] restriction of range derived by Kelly [1923, p. 316] using the method of logarithmic differentials. The empirically derived nomographs for the construction of confidence intervals for the correlation corrected for Case 2 [Case A] restriction of range presented by Gulliksen and Hopkins [1976] provide support for the use of Equation 4.26 in as far as the results obtained from Equation 4.26 and the standard error implied by the nomographs agree [Bobko & Rieck, 1980].

Figures 4.17 - 4.18 depict the reaction of  $\sigma[\rho^{\wedge}_c]$  to changes in  $K$ ,  $\rho[x,y]$  and  $n$ . Figure 4.19 and Figure 4.20 present an alternative perspective on the behaviour of  $\sigma[\rho^{\wedge}_c]$  by rotating the three dimensional space through 60° towards the  $\rho[x,y]$  axis. Although it is rather difficult to precisely capture the idiosyncrasies in the behaviour of  $\sigma[\rho^{\wedge}_c]$  in words, Figures 4.17 - 4.20 suggest that:

- $\sigma[\rho^{\wedge}_c]$  increases as  $\rho[x,y]$  decreases;
- $\sigma[\rho^{\wedge}_c]$  increases relatively sharply as  $K$  increases at low values of  $\rho[x,y]$ , but decreases slowly as  $K$  increases at high values of  $\rho[x,y]$ ;
- $\sigma[\rho^{\wedge}_c]$  increases as sample size decreases.

- The inflection point in the negative relationship between  $\sigma[\hat{\rho}_c]$  and  $\rho[x,y]$  seems to gradually shift upwards on the  $\rho[x,y]$  axis as  $K$  increases.

Figures 4.21 - 4.22 depict the behavior of the standard error ratio  $V$  to changes in  $K$ ,  $\rho[x,y]$  and  $n$ . Figure 4.23 and Figure 4.24 present an alternative perspective on the behaviour of  $V$  by rotating the three dimensional space through  $60^\circ$  towards the  $\rho[x,y]$  axis. The reaction of  $V$  to changes in  $K$ ,  $\rho[x,y]$  and  $n$  seems to mirror the behaviour of  $\sigma[\hat{\rho}_c]$  as described above. A noteworthy quality of the behaviour of  $V$  is its habit to decrease below unity in a relatively small region of the  $\rho[x,y]$ ,  $K$  plane characterised by high  $\rho[x,y]$  and  $K$  values. When estimating  $\sigma[\hat{\rho}]$  through Equation 4.13,  $\sigma[\hat{\rho}_c]$  estimated through Equation 4.35 will thus exceed the former at small selection ratios and high restricted [i.e. uncorrected] correlations.

Using a Monte Carlo simulation, Mendoza, Hart and Powell [1991] evaluated a bootstrap procedure for setting a confidence interval on the correlation coefficient corrected for restriction of range on the predictor. In their investigation of the behaviour of the corrected correlation coefficient Mendoza, Hart and Powell [1991] considered only incomplete truncation on the predictor  $X$  although they used the usual Pearson correction formula. Incomplete truncation on the predictor refers to a situation where observation of individuals over the full range of predictor and criterion scores is possible but the probability of not being able to observe a criterion performance is related to predictor score [Olson & Becker, 1983]. Mendoza, Hart and Powell [1991, p. 268] summarise the findings of their investigation as follow:

The bootstrap interval was shown to be accurate under incomplete truncation, lack of symmetry, and lack of homoscedasticity. Also the standard corrected correlation  $r_c$  was robust to incomplete truncation and homogeneity when  $\rho$  was 0,50 or 0,30. On the other hand when either the population correlation or sample size was small, then  $r_c$  had the tendency to overestimate, even under distributions where homogeneity was not violated [normal and mixed]. Low truncation improved the precision of the confidence intervals and the accuracy of  $r_c$  in all of the distributions, whereas high truncation had a detrimental effect on the confidence intervals and the  $r_c$ . The effect was especially a problem when  $\rho$  was small. However, most or all of this effect seemed to have been due to sample size. In closing, we feel comfortable in recommending the use of the bootstrap interval when the sample size is at least 50, especially if  $\rho$  is believed to be moderate or large. We also recommend that the number of bootstrap iterations should be at least 2000.

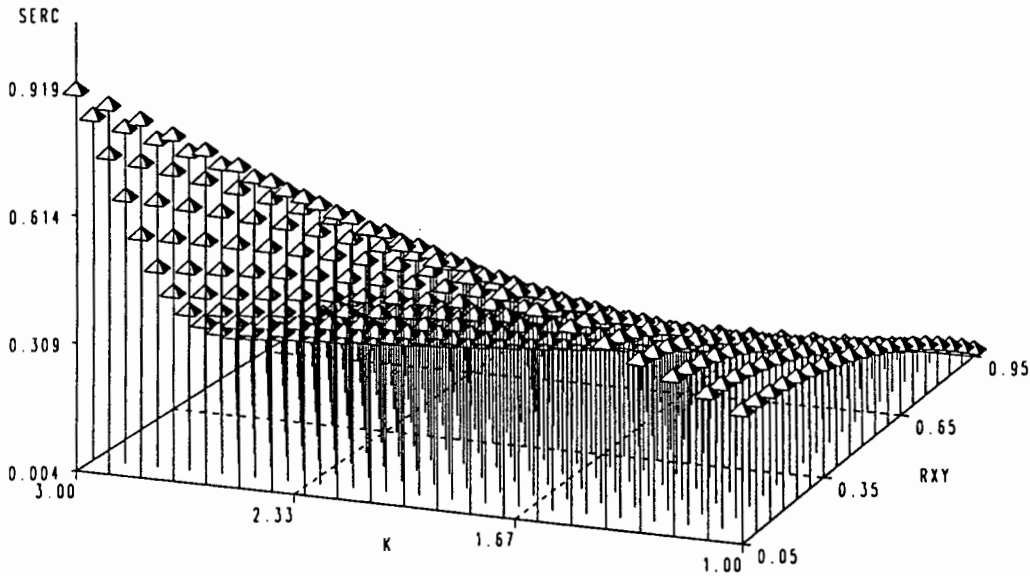


Figure 4.17: The standard error of the validity coefficient corrected for Case 2 [Case A] restriction of range as a function of K and  $\rho[x,y]$  for n fixed at 10

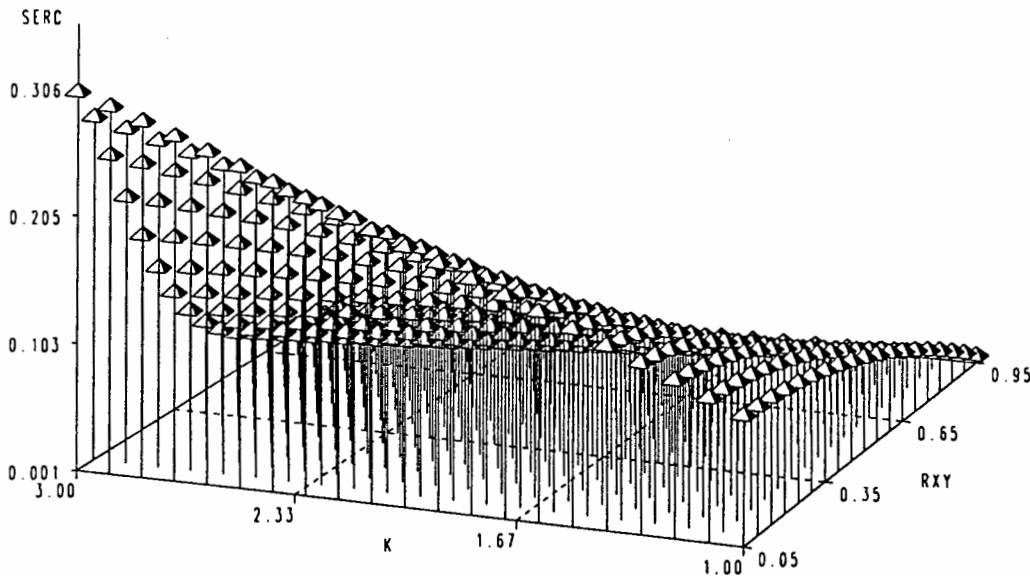


Figure 4.18: The standard error of the validity coefficient corrected for Case 2 [Case A] restriction of range as a function of K and  $\rho[x,y]$  for n fixed at 90

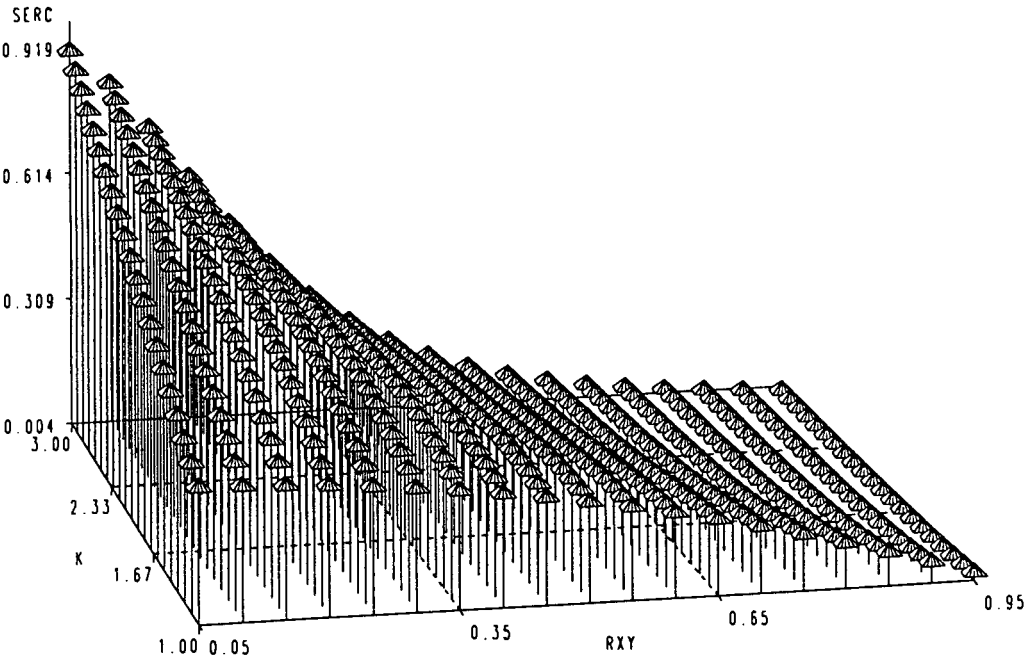


Figure 4.19: The standard error of the validity coefficient corrected for Case 2 [Case A] restriction of range as a function of K and  $\rho[x,y]$  [rotated 60°] for n fixed at 10

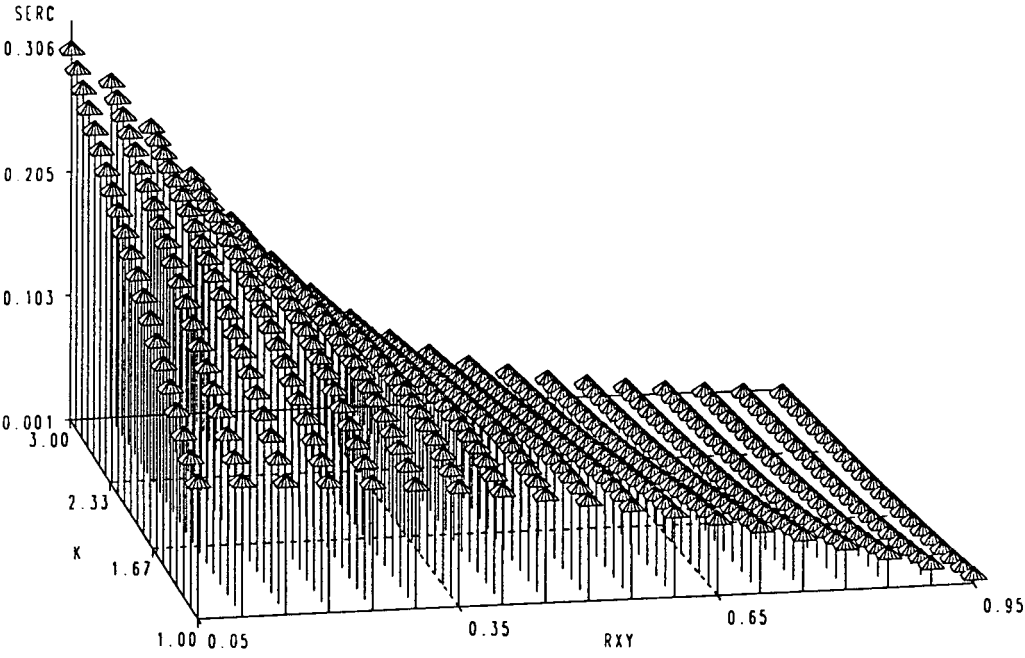


Figure 4.20: The standard error of the validity coefficient corrected for Case 2 [Case A] restriction of range as a function of K and  $\rho[x,y]$  [rotated 60°] for n fixed at 90

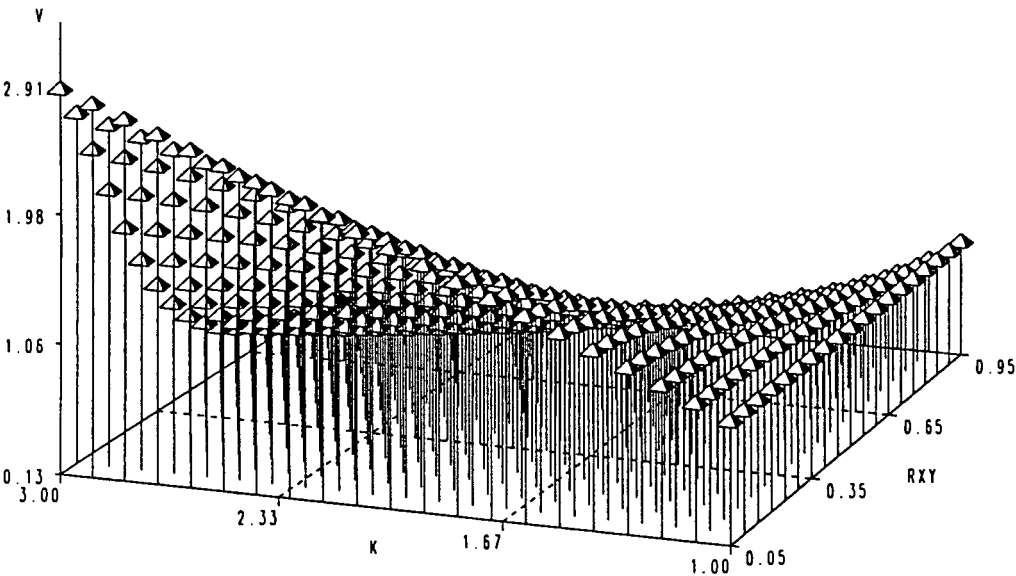


Figure 4.21:  $V = \sigma[\rho^*_c]/\sigma[\rho^*]$  as a function of  $K$  and  $\rho[x,y]$  for  $n$  fixed at 10

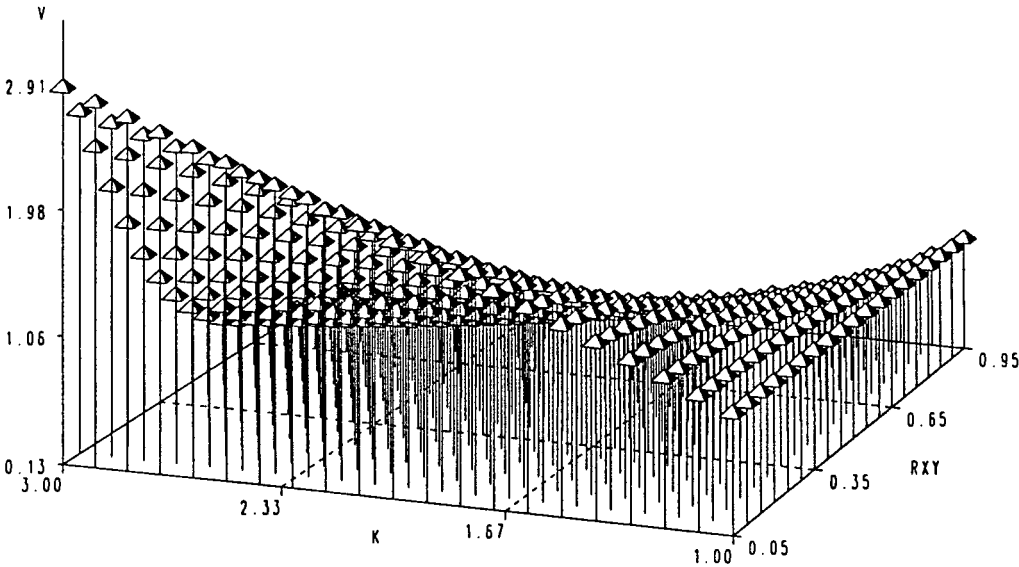


Figure 4.22:  $V = \sigma[\rho^*_c]/\sigma[\rho^*]$  as a function of  $K$  and  $\rho[x,y]$  for  $n$  fixed at 90



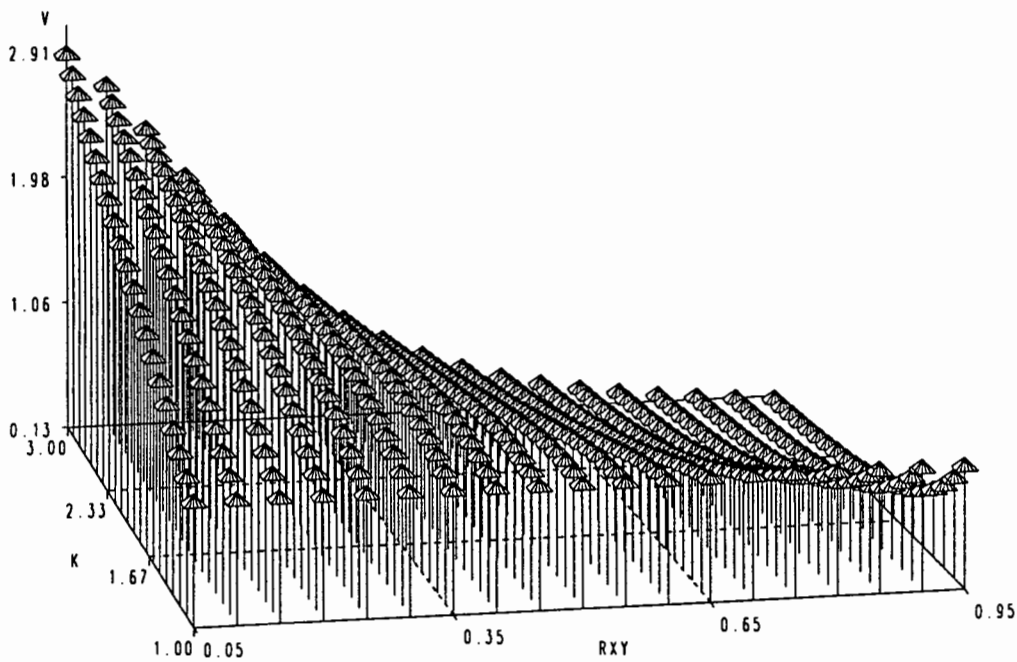


Figure 4.23:  $V = \sigma[\rho^{\wedge}_c]/\sigma[\rho^{\wedge}]$  as a function of  $K$  and  $\rho[x,y]$  [rotated through  $60^\circ$ ] for  $n$  fixed at 10

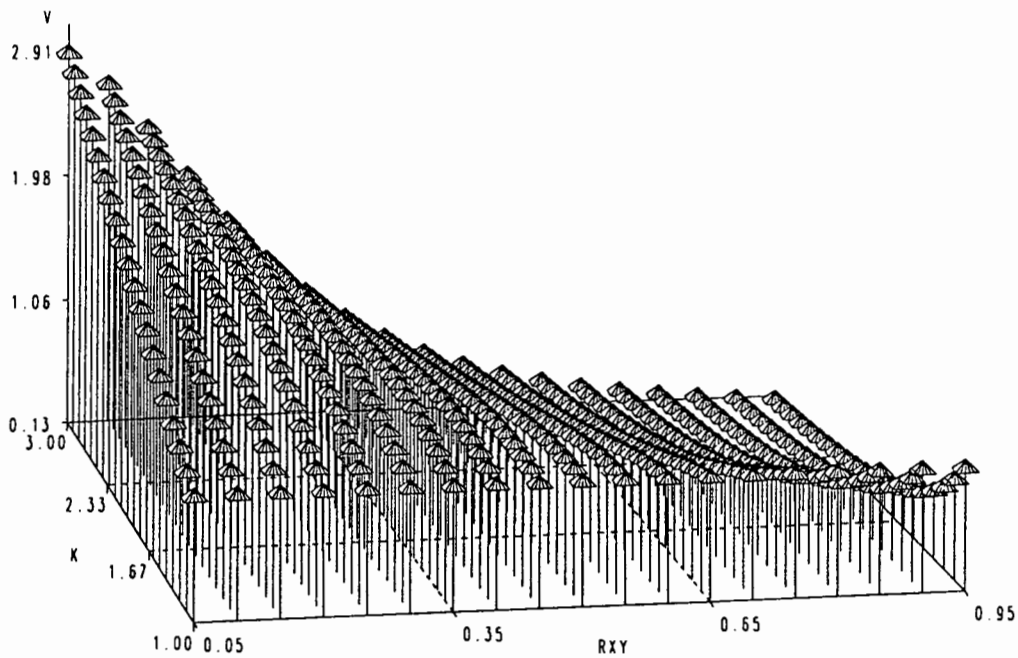


Figure 4.24:  $V = \sigma[\rho^{\wedge}_c]/\sigma[\rho^{\wedge}]$  as a function of  $K$  and  $\rho[x,y]$  [rotated through  $60^\circ$ ] for  $n$  fixed at 90



#### 4.6.6 Pearson Correlation Coefficient Corrected For Case C Restriction Of Range

Allen and Dunbar [1990] propose two large sample estimates of the sampling variance of a correlation corrected for indirect [Case 3[i] [Case C[i]]] selection. Both expressions were developed using the delta method. Following Kelley [1923], and Bobko and Rieck [1980], Allen and Dunbar [1990] assume the variance ratio  $\sigma^2[Z]/\sigma^2[z]$  to be fixed. Both standard error expressions can be stated as Equation 4.36 [Allan & Dunbar, 1990, p. 85].

$$\sqrt{\sigma^2[\rho^*[X,Y]]} = \left\{ \begin{aligned} &\rho'^2[X,Y]\sigma^2[\rho^*[x,y]] + \\ &\rho'^2[X,Z]\sigma^2[\rho^*[x,z]] + \\ &\rho'^2[Y,Z]\sigma^2[\rho^*[y,z]] + \\ &2\rho'[X,Y]\rho'[X,Z]\sigma[\rho^*[x,y],\rho^*[x,z]] + \\ &2\rho'[X,Y]\rho'[Y,Z]\sigma[\rho^*[x,y],\rho^*[y,z]] + \\ &2\rho'[X,Z]\rho'[Y,Z]\sigma[\rho^*[x,z],\rho^*[y,z]] \end{aligned} \right\}^{1/2} \quad 4.36$$

Where:

$\rho'[X,Y]$  indicates the partial derivative of the adjusted correlation  $\rho[X,Y]$  with respect to the unadjusted correlation  $\rho[x,y]$ ; specifically

$$\rho'[X,Y] = (1 + W\rho^2[x,z])^{-1/2} (1 + W\rho^2[y,z])^{-1/2}$$

$$\rho'[X,Z] = W(1 + W\rho^2[x,z])^{-3/2} (1 + W\rho^2[y,z])^{-1/2} (\rho[y,z] - \rho[x,y]\rho[x,z])$$

$$\rho'[Y,Z] = W(1 + W\rho^2[x,z])^{-1/2} (1 + W\rho^2[y,z])^{-3/2} (\rho[x,z] - \rho[x,y]\rho[y,z]); \text{ and}$$

$$W = (\sigma^2[Z]/\sigma^2[z]) - 1$$

The two large sample estimates of the sampling variance of a correlation corrected for indirect [Case 3[i] [Case C[i]]] selection proposed by Allen & Dunbar [1990] differ in terms of the equations chosen for the estimation of the asymptotic variance and covariance components. The difference fundamentally lies in the assumptions on which these estimates are based. Allen and Dunbar [1990, p. 86] distinguish between two sets of assumptions, namely:

- set A; "error/n to second and third powers is negligible, linearity of regression, and mesokurtosis of the joint distribution of explicit and implicit variables; and
- set B; error/n to second and third powers is negligible and linearity of regression."

Only the variance and covariance equations based on the more stringent assumption set A are presented below.

$$\sigma[\rho^{\wedge}[x,y],\rho^{\wedge}[x,z]]=(1/n)\{\rho[y,z](1-\rho^2[x,z]-\rho^2[x,y])^{-1/2}\rho[x,y]\rho[x,z](1-\rho^2[y,z]-\rho^2[x,y]-\rho^2[x,z])\}$$

$$\sigma[\rho^{\wedge}[x,y],\rho^{\wedge}[y,z]]=(1/n)\{\rho[x,z](1-\rho^2[x,y]-\rho^2[y,z])^{-1/2}\rho[x,y]\rho[y,z](1-\rho^2[x,z]-\rho^2[x,y]-\rho^2[y,z])\}$$

$$\sigma[\rho^{\wedge}[x,z],\rho^{\wedge}[y,z]]=(1/n)\{\rho[x,y](1-\rho^2[x,z]-\rho^2[y,z])^{-1/2}\rho[x,z]\rho[y,z](1-\rho^2[x,y]-\rho^2[x,z]-\rho^2[y,z])\}$$

$$\sigma^2[\rho^{\wedge}[x,z]] = (1 - \rho^2[x,z])^2/n$$

$$\sigma^2[\rho^{\wedge}[y,z]] = (1 - \rho^2[y,z])^2/n$$

$$\sigma^2[\rho^{\wedge}[x,y]] = (1 - \rho^2[x,y])^2/n$$

Although the assumptions of these estimates are most likely not precisely met in actual Case C selection settings, some of the quantities in the alternative variance and covariance expressions are to such an extent difficult to estimate that the easier expressions are rather given precedence. Allen and Dunbar [1990, p. 91] summarise their position on the different possible approaches to obtain an estimate of the standard error for the correlation corrected for Case 3[i] [Case C[i]] restriction of range as follow:

Comparison of the alternative estimates of standard errors led to one dominant finding: For settings permitting some degree of confidence in the Pearson-Lawley adjustments, the SE estimates provided very similar assessments of the degree of sampling error. Given these results, the use of the simpler of the two large-sample estimates is suggested for both the two and three variable cases when sample size is large. Although the assumptions for this estimate may not be strictly true, this approximation appeared to offer a reasonable estimate of the SE of an adjusted correlation without resorting to the computer-intensive approach of the bootstrap estimate or the complexity of the large-sample estimate with less stringent assumptions.

The reaction of  $\sigma[\rho^{\wedge}_c]$  to changes in  $\rho[x,y]$ ,  $K$ ,  $\rho[x,z]$ ,  $\rho[y,z]$  and  $n$  is graphically portrayed in Figures 4.25 - 4.38. Figure 4.37 and Figure 4.38 provide alternative perspectives on the behavior of  $\sigma[\rho^{\wedge}_c]$  by rotating the respective spaces through  $60^\circ$  towards the  $\rho[x,z]$  axis. Figures 4.25 - 4.38 collectively clearly indicate the complexity of  $\sigma[\rho^{\wedge}_c]$ 's behaviour under changes in the relevant parameters.

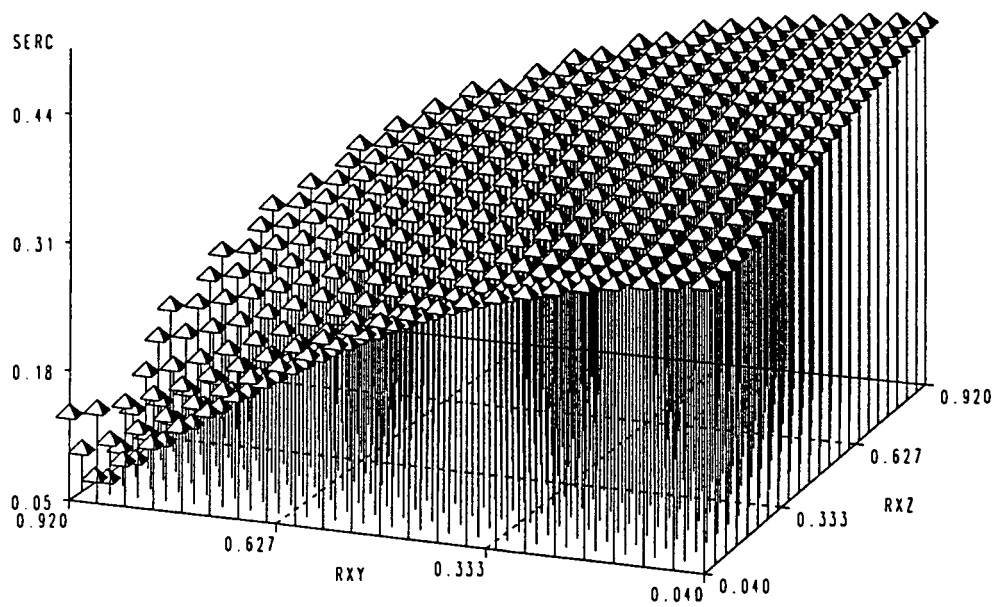


Figure 4.25: The standard error of the validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of  $\rho[x,y]$  and  $\rho[x,z]$  for  $n$  fixed at 10,  $K$  fixed at 2 and  $\rho[y,z]$  fixed at 0.15

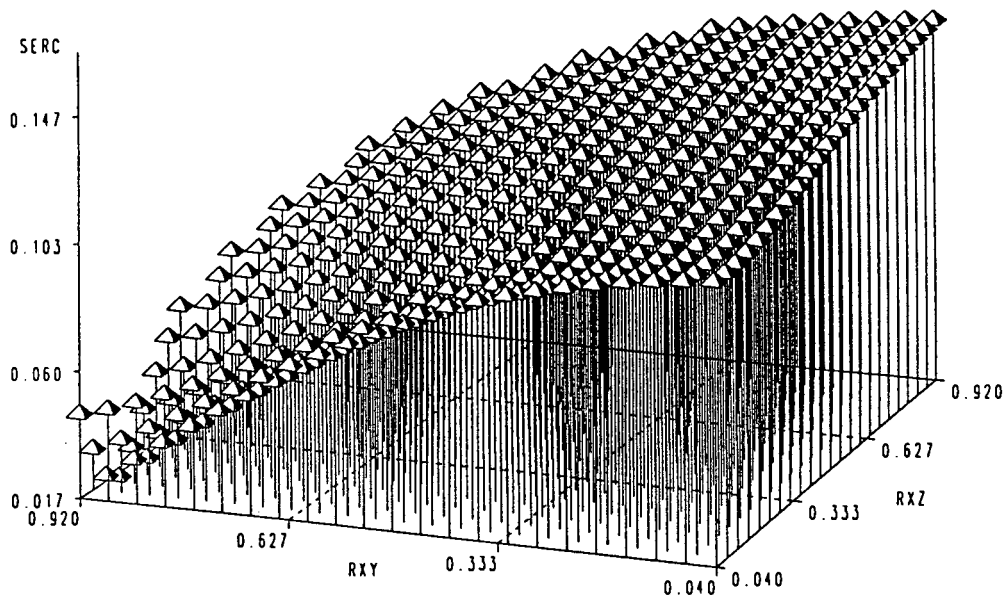


Figure 4.26: The standard error of the validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of  $\rho[x,y]$  and  $\rho[x,z]$  for  $n$  fixed at 90,  $K$  fixed at 2 and  $\rho[y,z]$  fixed at 0.15

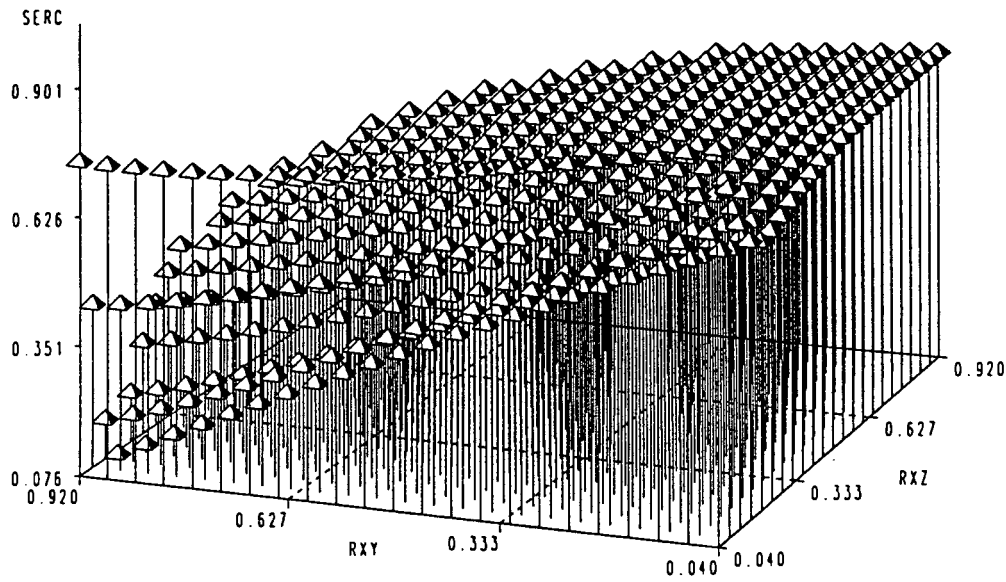


Figure 4.27: The standard error of the validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of  $\rho[x,y]$  and  $\rho[x,z]$  for  $n$  fixed at 10,  $K$  fixed at 5 and  $\rho[y,z]$  fixed at 0.15

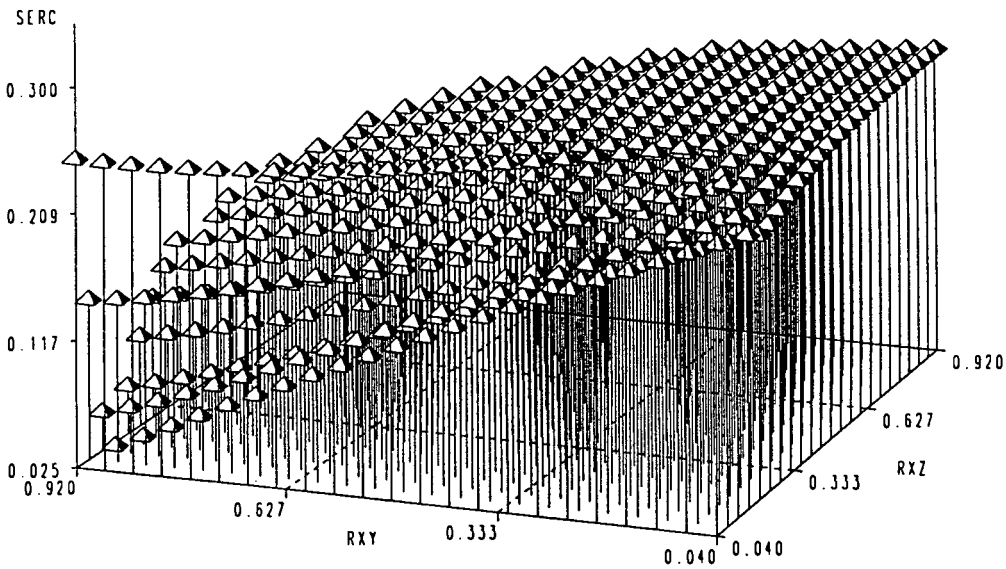


Figure 4.28: The standard error of the validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of  $\rho[x,y]$  and  $\rho[x,z]$  for  $n$  fixed at 90,  $K$  fixed at 2 and  $\rho[y,z]$  fixed at 0.15

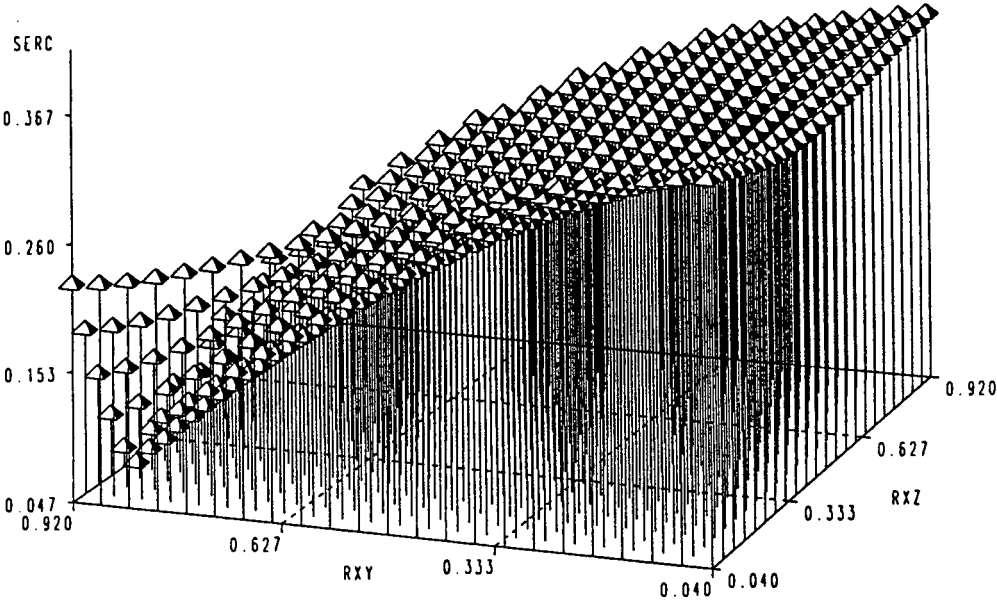


Figure 4.29: The standard error of the validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of  $\rho[x,y]$  and  $\rho[x,z]$  for  $n$  fixed at 10,  $K$  fixed at 2 and  $\rho[y,z]$  fixed at 0.25

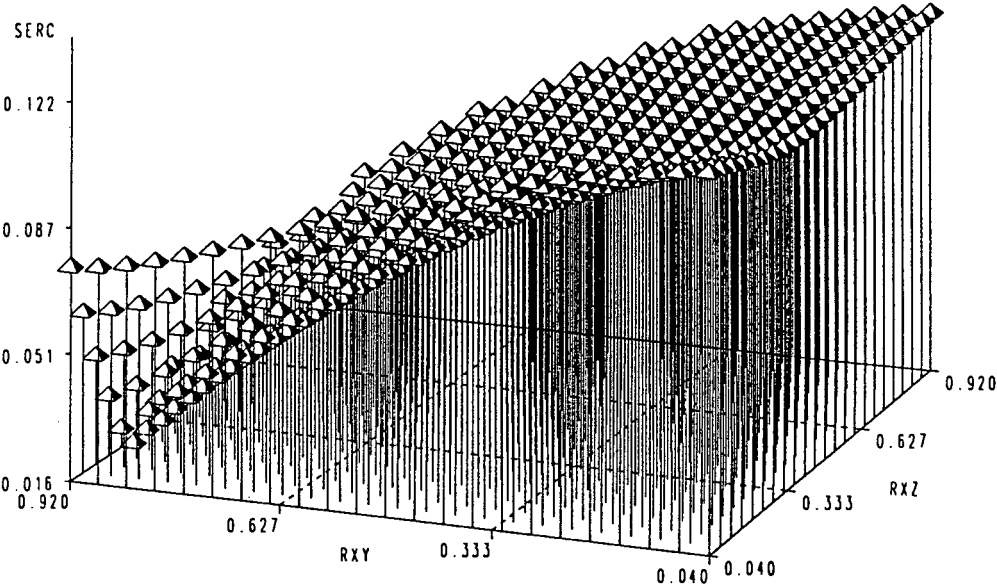


Figure 4.30: The standard error of the validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of  $\rho[x,y]$  and  $\rho[x,z]$  for  $n$  fixed at 90,  $K$  fixed at 2 and  $\rho[y,z]$  fixed at 0.25

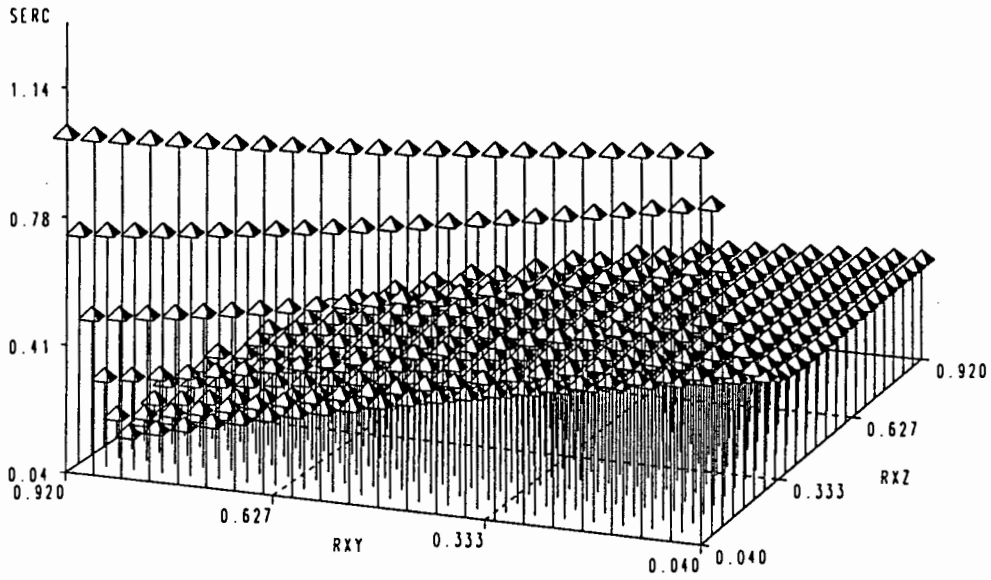


Figure 4.31: The standard error of the validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of  $\rho[x,y]$  and  $\rho[x,z]$  for  $n$  fixed at 10,  $K$  fixed at 5 and  $\rho[y,z]$  fixed at 0.25

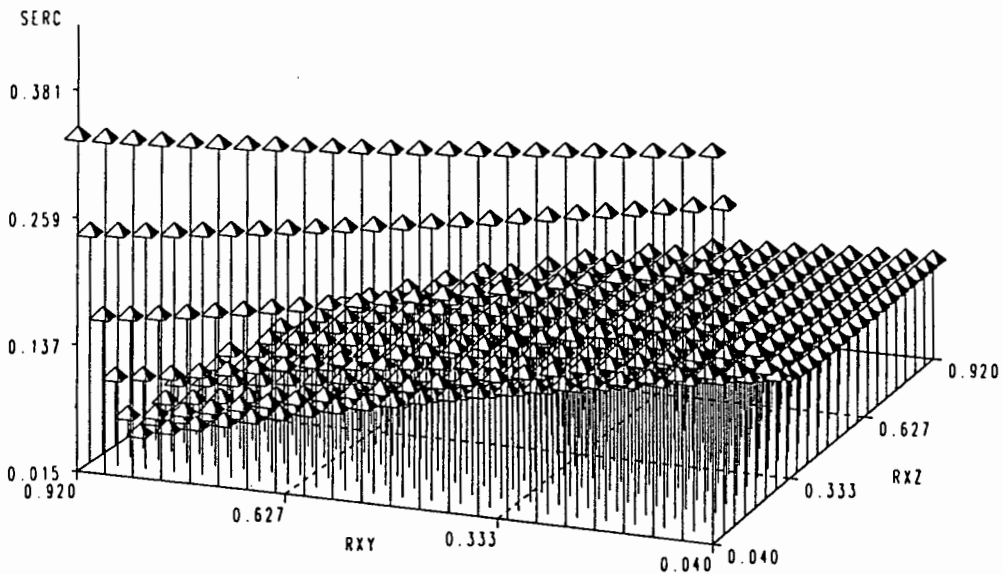


Figure 4.32: The standard error of the validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of  $\rho[x,y]$  and  $\rho[x,z]$  for  $n$  fixed at 90,  $K$  fixed at 5 and  $\rho[y,z]$  fixed at 0.25



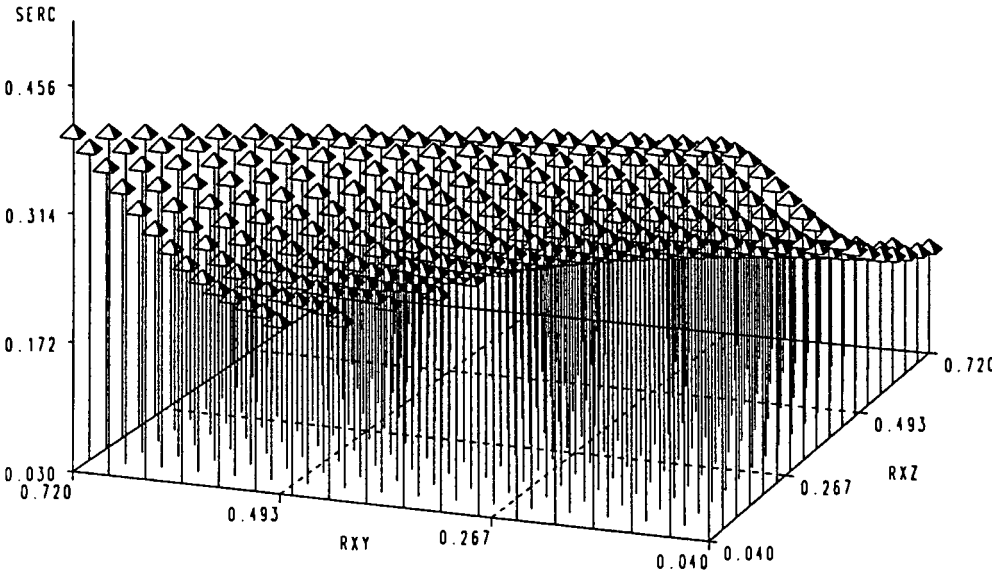


Figure 4.33: The standard error of the validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of  $\rho[x,y]$  and  $\rho[x,z]$  for  $n$  fixed at 10,  $K$  fixed at 2 and  $\rho[y,z]$  fixed at 0.65

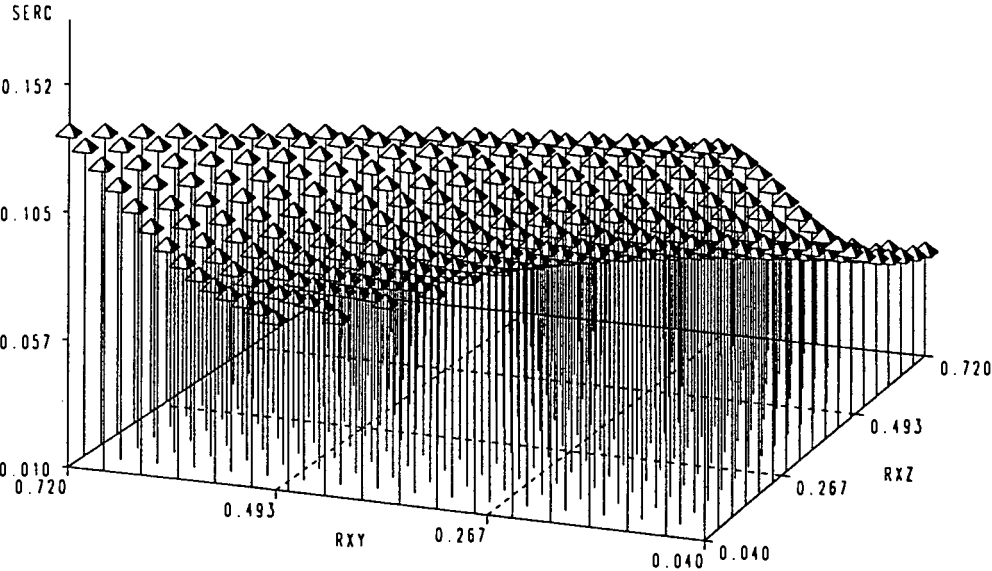


Figure 4.34: The standard error of the validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of  $\rho[x,y]$  and  $\rho[x,z]$  for  $n$  fixed at 90,  $K$  fixed at 2 and  $\rho[y,z]$  fixed at 0.65

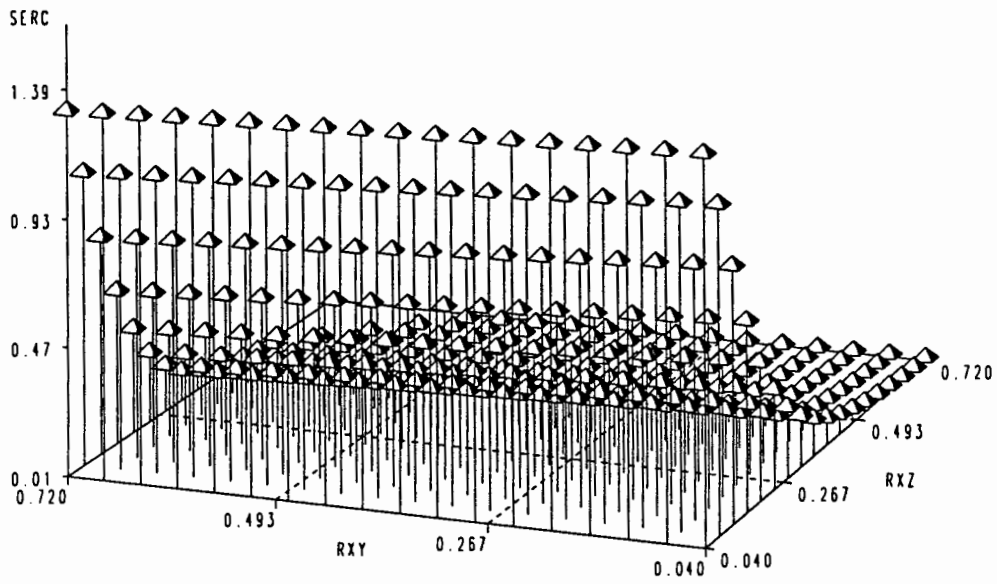


Figure 4.35: The standard error of the validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of  $\rho[x,y]$  and  $\rho[x,z]$  for  $n$  fixed at 10,  $K$  fixed at 5 and  $\rho[y,z]$  fixed at 0.65

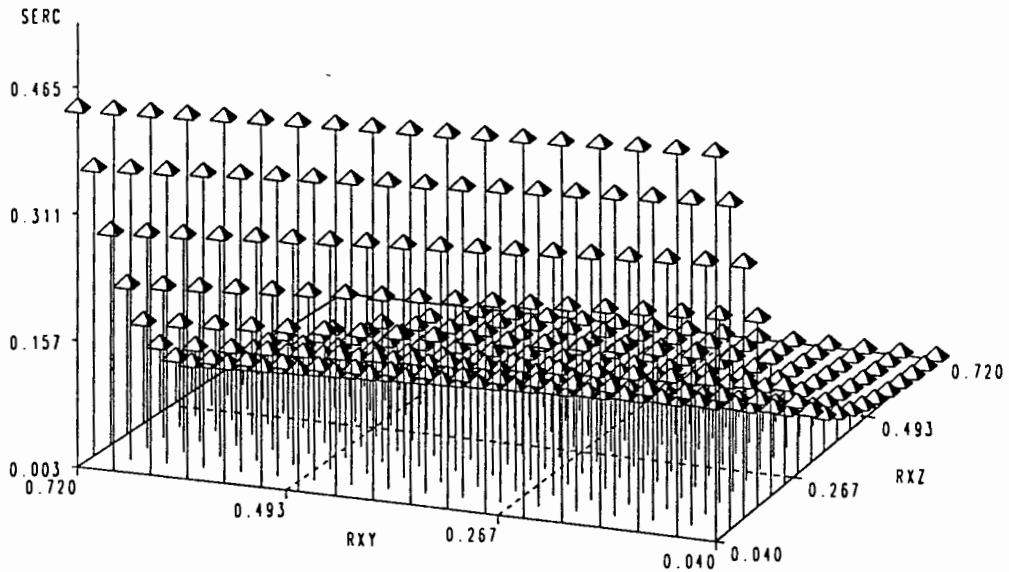


Figure 4.36: The standard error of the validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of  $\rho[x,y]$  and  $\rho[x,z]$  for  $n$  fixed at 90,  $K$  fixed at 5 and  $\rho[y,z]$  fixed at 0.65

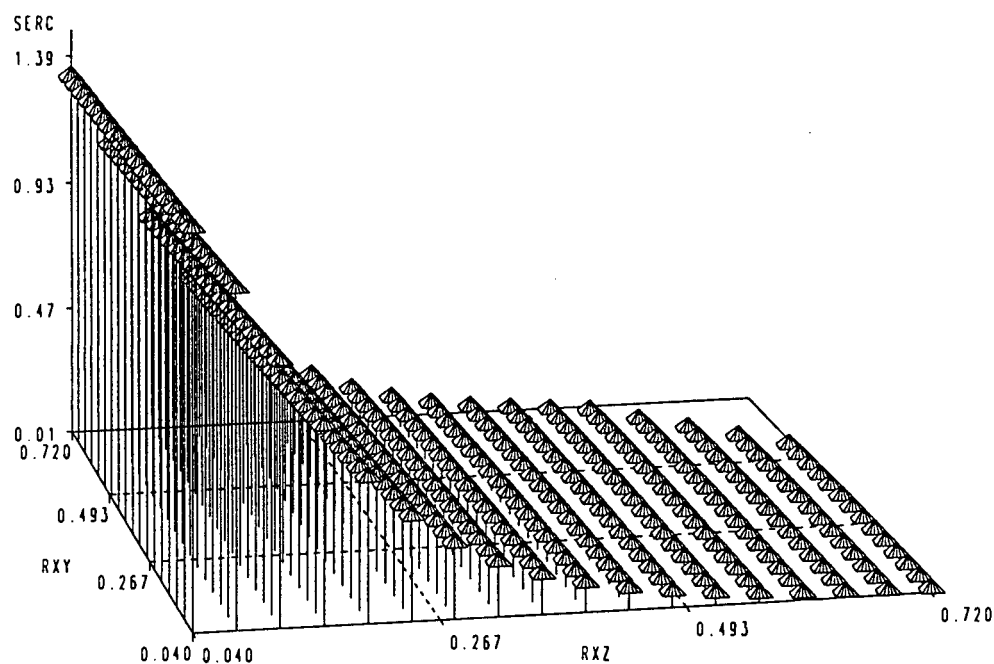


Figure 4.37: The standard error of the validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of  $\rho[x,y]$  and  $\rho[x,z]$  [rotated through  $60^\circ$ ] for  $n$  fixed at 10,  $K$  fixed at 5 and  $\rho[y,z]$  fixed at 0.65

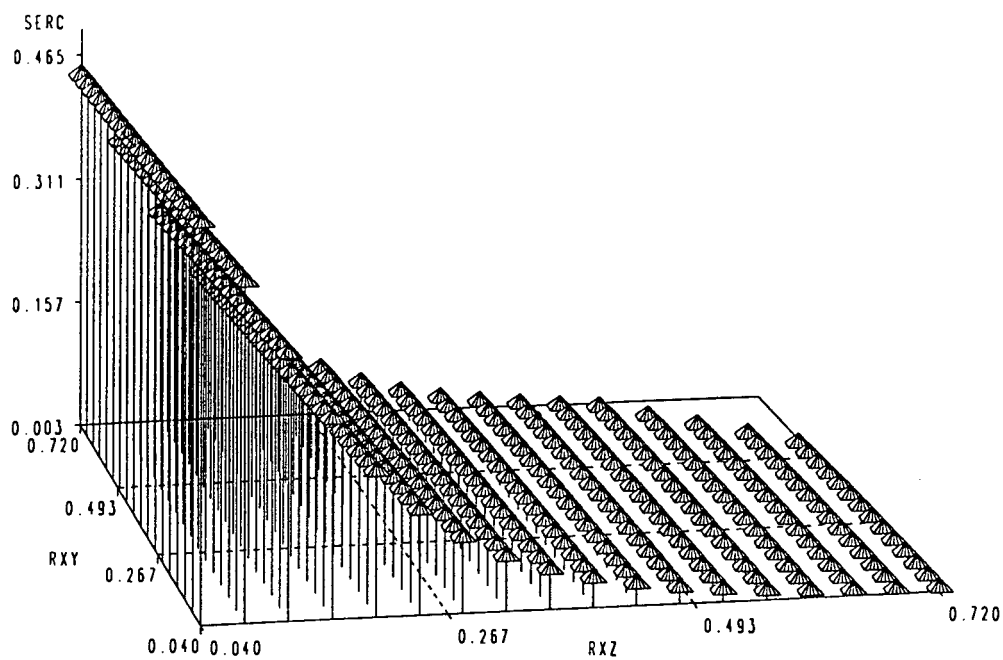


Figure 4.38: The standard error of the validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range as a function of  $\rho[x,y]$  and  $\rho[x,z]$  [rotated through  $60^\circ$ ] for  $n$  fixed at 90,  $K$  fixed at 5 and  $\rho[y,z]$  fixed at 0.65

Although the complexity tends to frustrate attempts to capture  $\sigma[\rho^{\wedge}_c]$ 's behaviour in words, Figures 4.25 - 4.38 seem to suggest that  $\sigma[\rho^{\wedge}_c]$ :

- decreases as  $n$  increases;
- increases as  $K$  increases [i.e. the selection ratio decreases];
- increases as  $\rho[x,y]$  decreases, but with the rate of increase dependent on  $\rho[x,z]$ ,  $\rho[y,z]$  and  $K$ ;
- the rate of increase accelerates as  $\rho[x,z]$  increases and then decreases again as  $\rho[x,z]$  increases further;
- the point of maximum slope [relative to  $\rho[x,y]$ ] on the  $\rho[x,z]$  axis shifts upwards on the latter axis as  $\rho[y,z]$  increases;
- the rate of increase decelerates as  $K$  increases.

The reaction of the standard error ratio,  $V = \sigma[\rho^{\wedge}_c]/\sigma[\rho^{\wedge}]$ , to changes in  $\rho[x,y]$ ,  $K$ ,  $\rho[x,z]$  and  $\rho[y,z]$  is graphically portrayed in Figures 4.39 - 4.50. Figures 4.39 - 4.50 indicate that the standard error of the uncorrected correlation coefficient can, under certain conditions, exceed the standard error of the corrected coefficient [i.e.  $V < 1$ ]. These conditions seem to depend on all parameters, except sample size. The conditions favoring  $V < 1$  seem to be smaller selection ratios [i.e. bigger  $K$ ], higher correlations between the selection variable  $Z$  and the criterion and predictor variables respectively [in the selected group] and lower  $\rho[x,y]$  values. The region of maximum difference in the magnitude of the standard errors seems to occur at high levels of  $\rho[x,y]$  and a reciprocal combination of  $\rho[x,z]$  and  $\rho[y,z]$ .

#### 4.6.7 Pearson Correlation Coefficient Double Corrected For Restriction Of Range And Criterion Unreliability

Applying a Taylor series approximation to Equation 3.10 [or equivalently, Equation 3.13], Bobko [1983] shows that the variance of the sampling distribution of the double corrected correlation coefficient [i. e. the square of the standard error of the double corrected correlation] can be expressed as Equation 4.37.

$$\sigma^2[\rho^{\wedge}[[X,T_Y]]] = n^{-1}(\rho_{tty})^{-1}K^2D^3\{(1-\rho^2[x,y])^2 + (1/4)\rho^2[x,y]\rho^{-2}_{tty}(1-\rho^2_{tty})^2 - \{\rho[x,y]\rho^{-1}_{tty}\}\{\rho[x,y](1-\rho^2[x,y]-\rho^2_{tty}) - (1/2)\rho[x,y]\rho_{tty}(1-2\rho^2[X,Y]-\rho^2_{tty})\}\} \quad 4.37$$

Where:

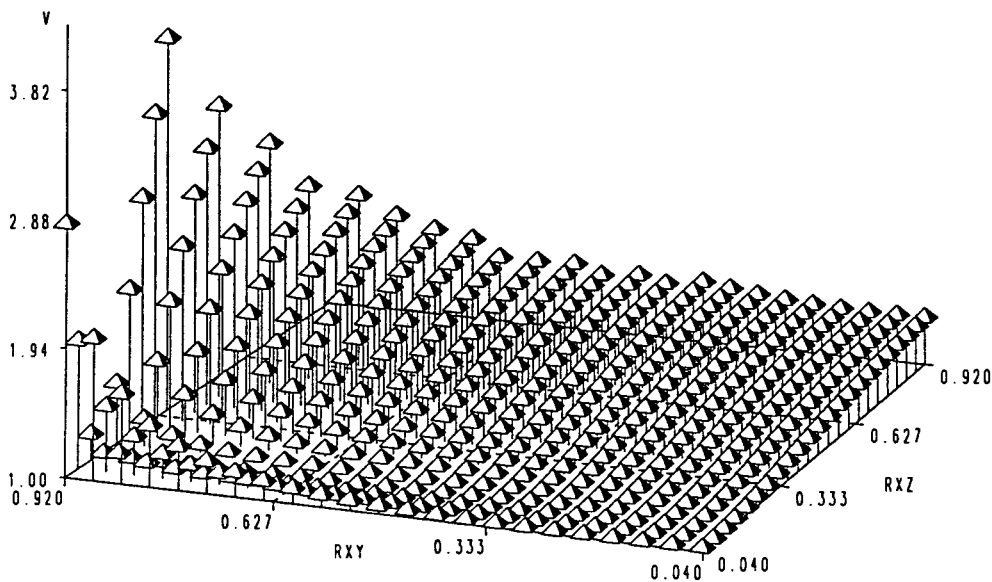


Figure 4.39:  $V = \sigma[\hat{\rho}] / \sigma[\rho]$  as a function of  $\rho[x,y]$  and  $\rho[x,z]$  for Case 3[i] [Case C[i]] restriction of range; n is fixed at 10, K is fixed at 2 and  $\rho[y,z]$  is fixed at 0.15

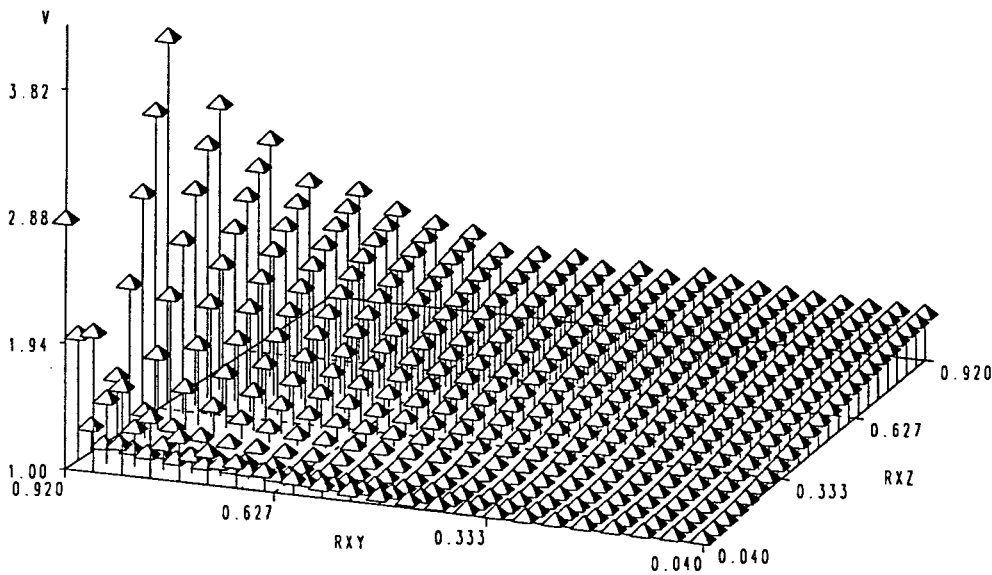


Figure 4.40:  $V = \sigma[\hat{\rho}] / \sigma[\rho]$  as a function of  $\rho[x,y]$  and  $\rho[x,z]$  for Case 3[i] [Case C[i]] restriction of range; n is fixed at 90, K is fixed at 2 and  $\rho[y,z]$  is fixed at 0.15

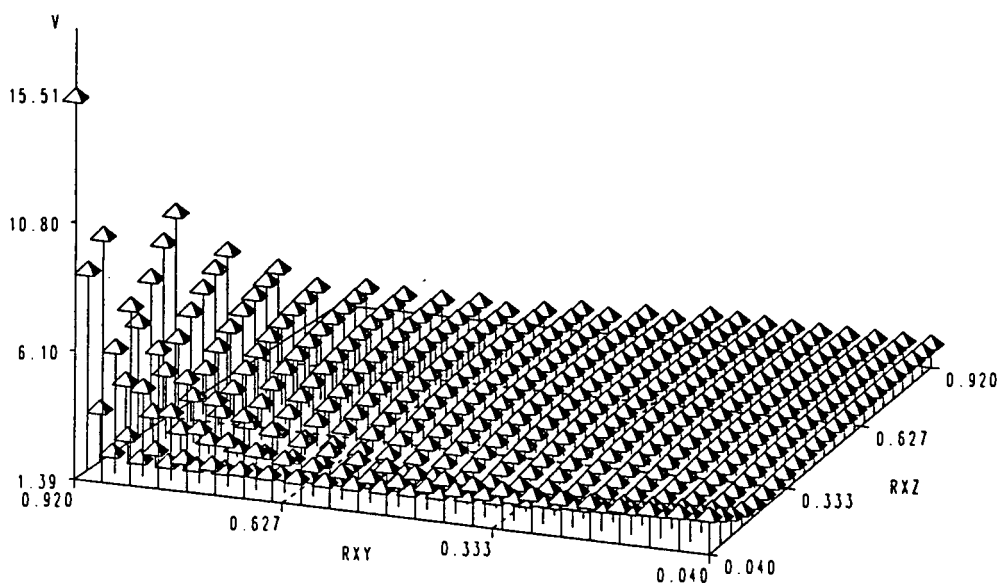


Figure 4.41:  $V = \sigma[\rho^*] / \sigma[\rho]$  as a function of  $\rho[x,y]$  and  $\rho[x,z]$  for Case 3[i] [Case C[i]] restriction of range; n is fixed at 10, K is fixed at 5 and  $\rho[y,z]$  is fixed at 0.15

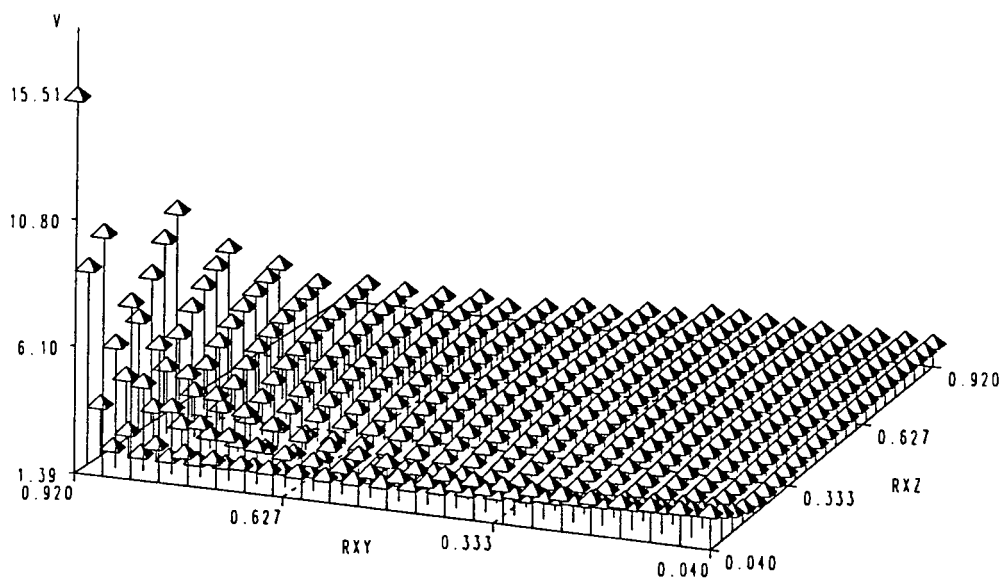


Figure 4.42:  $V = \sigma[\rho^*] / \sigma[\rho]$  as a function of  $\rho[x,y]$  and  $\rho[x,z]$  for Case 3[i] [Case C[i]] restriction of range; n is fixed at 90, K is fixed at 5 and  $\rho[y,z]$  is fixed at 0.15



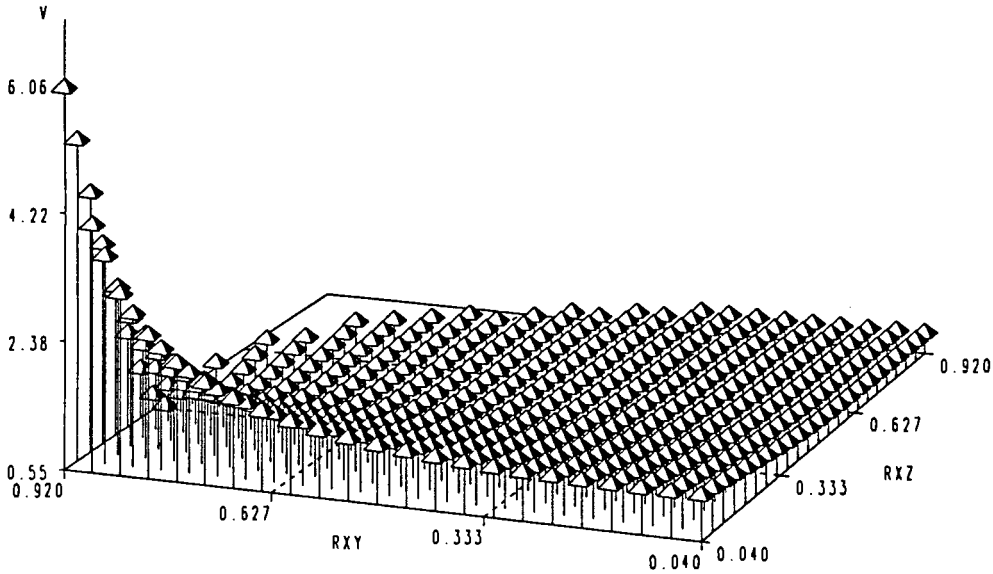


Figure 4.43:  $V = \sigma[\rho^*_c]/\sigma[\rho]$  as a function of  $\rho[x,y]$  and  $\rho[x,z]$  for Case 3[i] [Case C[i]] restriction of range; n is fixed at 10, K is fixed at 2 and  $\rho[y,z]$  is fixed at 0.35

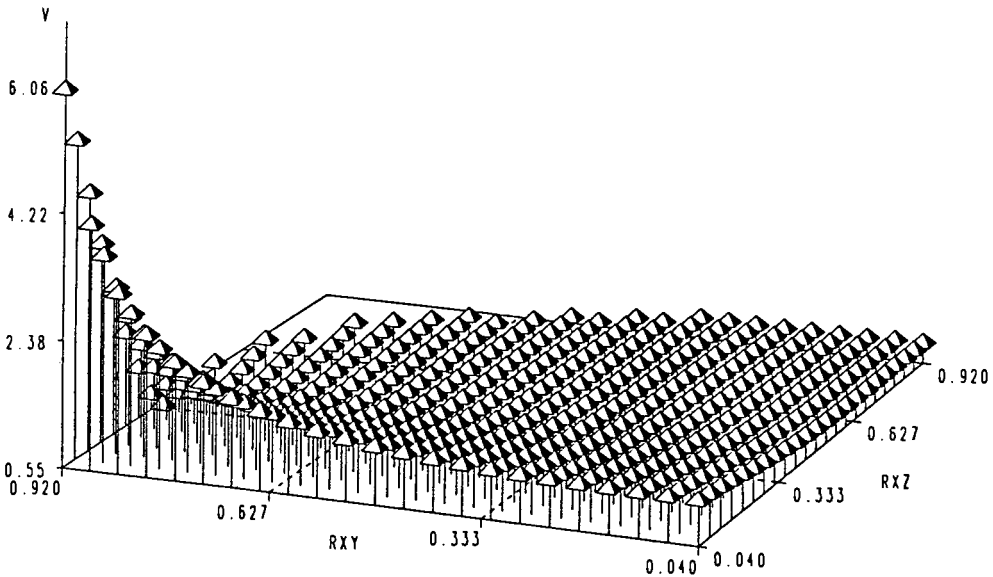


Figure 4.44:  $V = \sigma[\rho^*_c]/\sigma[\rho]$  as function of  $\rho[x,y]$  and  $\rho[x,z]$  for Case 3[i] [Case C[i]] restriction of range; n is fixed at 90, K is fixed at 2 and  $\rho[y,z]$  is fixed at 0.35

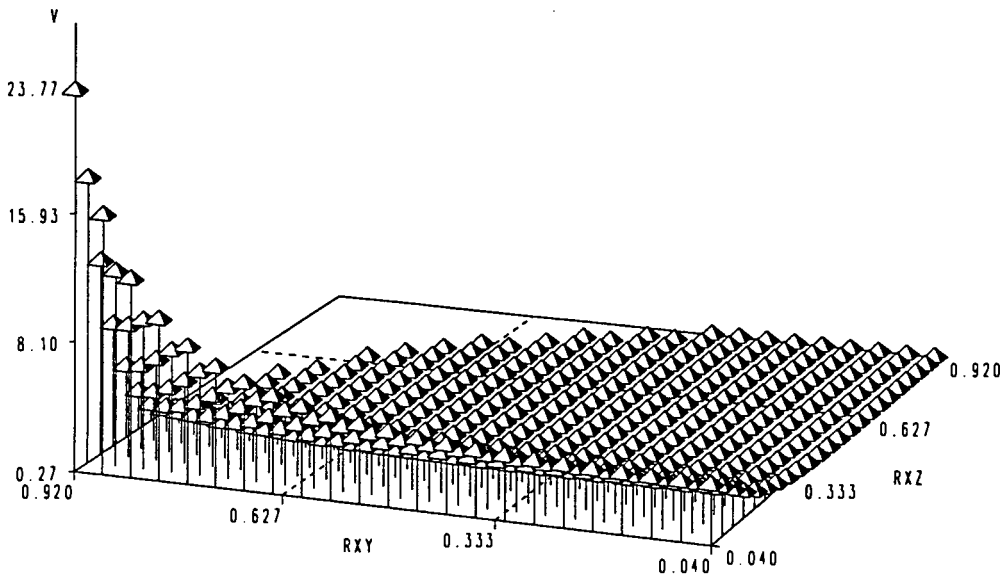


Figure 4.45:  $V = \sigma[\rho^{\wedge}_{\cdot}]/\sigma[\rho]$  as a function of  $\rho[x,y]$  and  $\rho[x,z]$  for Case 3[i] [Case C[i]] restriction of range; n is fixed at 10, K is fixed at 5 and  $\rho[y,z]$  is fixed at 0.35

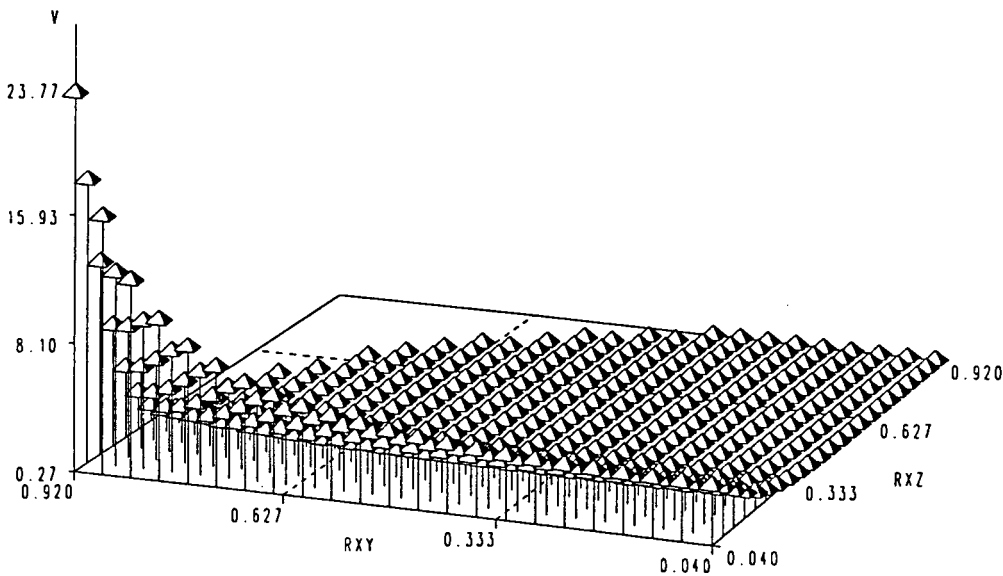


Figure 4.46:  $V = \sigma[\rho^{\wedge}_{\cdot}]/\sigma[\rho]$  as function of  $\rho[x,y]$  and  $\rho[x,z]$  for Case 3[i] [Case C[i]] restriction of range; n is fixed at 90, K is fixed at 5 and  $\rho[y,z]$  is fixed at 0.35

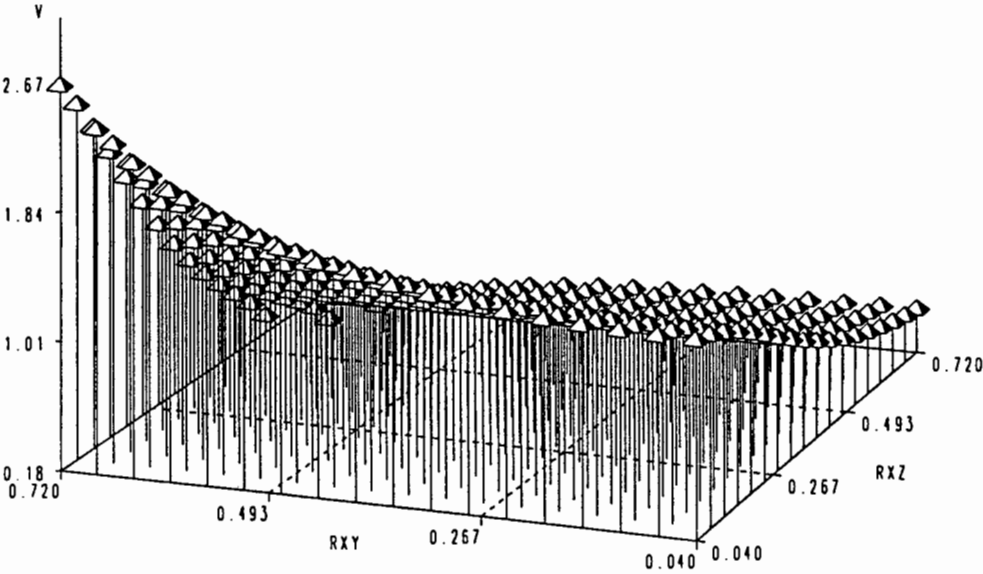


Figure 4.47:  $V = \sigma[\hat{\rho}] / \sigma[\rho]$  as a function of  $\rho[x,y]$  and  $\rho[x,z]$  Case 3[i] [Case C[i]] restriction of range; n is fixed at 10, K is fixed at 2 and  $\rho[y,z]$  is fixed at 0.65

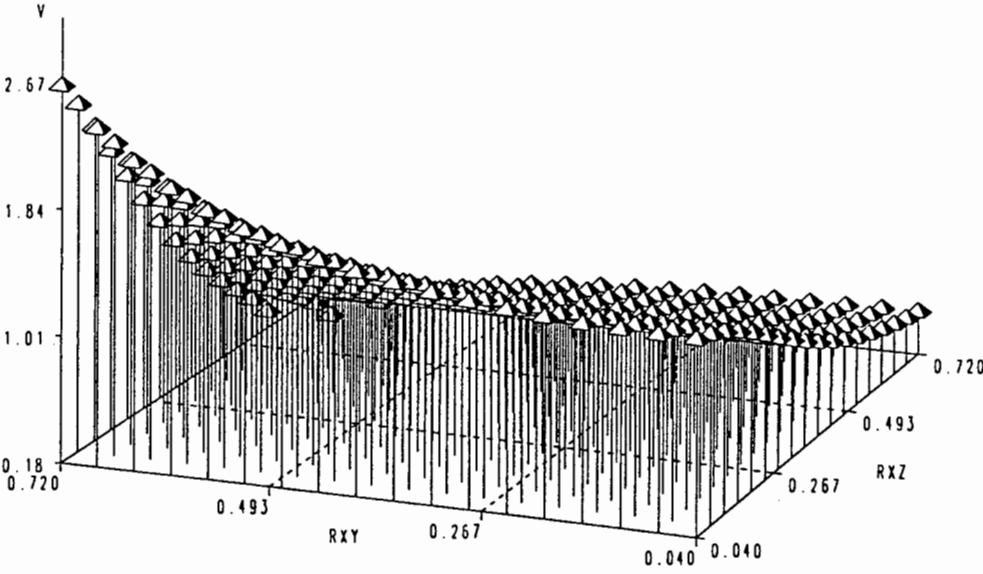


Figure 4.48:  $V = \sigma[\hat{\rho}] / \sigma[\rho]$  as a function of  $\rho[x,y]$  and  $\rho[x,z]$  for Case 3[i] [Case C[i]] restriction of range; n is fixed at 90, K is fixed at 2 and  $\rho[y,z]$  is fixed at 0.65

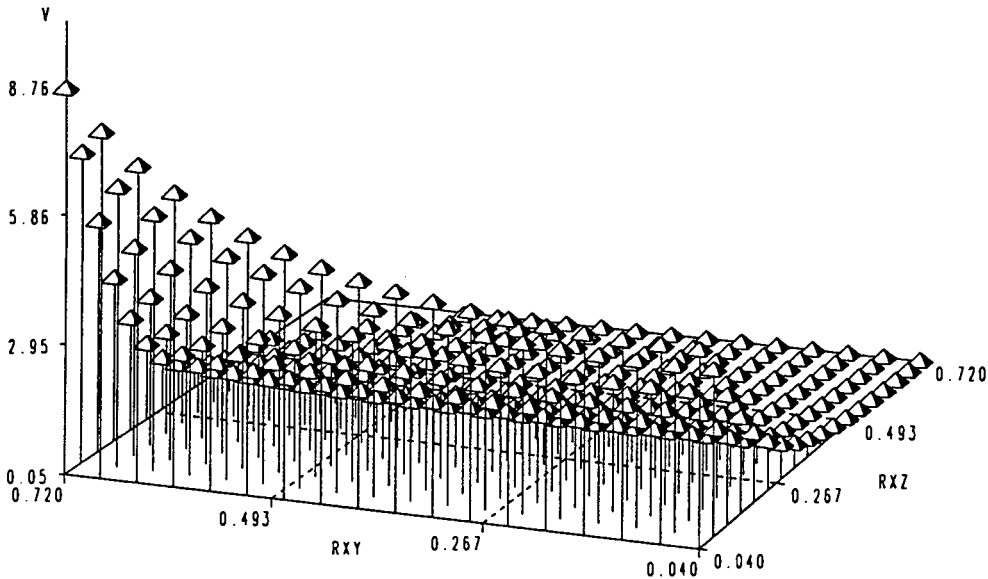


Figure 4.49:  $V = \sigma[\hat{\rho}] / \sigma[\rho]$  as a function of  $\rho[x,y]$  and  $\rho[x,z]$  for Case 3[i] [Case C[i]] restriction of range; n is fixed at 10, K is fixed at 5 and  $\rho[y,z]$  is fixed at 0.65

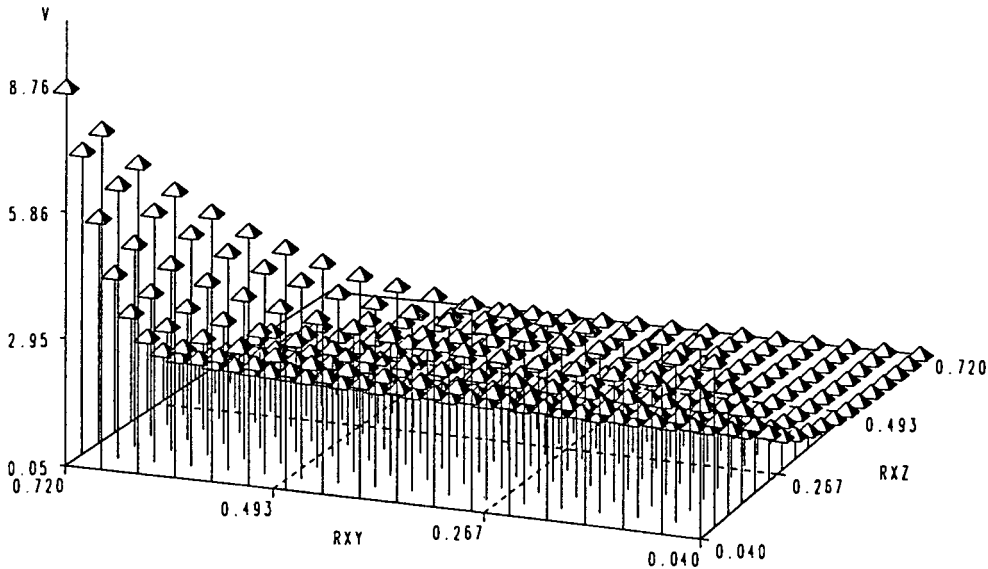


Figure 4.50:  $V = \sigma[\hat{\rho}] / \sigma[\rho]$  as a function of  $\rho[x,y]$  and  $\rho[x,z]$  for Case 3[i] [Case C[i]] restriction of range ; n is fixed at 90, K is fixed at 5 and  $\rho[y,z]$  is fixed at 0.65

$$D = (1 - \rho^2[X, Y] \rho^{-1}_{\text{tty}} + K^2 \rho^2[X, Y] \rho^{-1}_{\text{tty}}), \text{ and}$$

$$K = \sigma[X] / \sigma[x]$$

The expression reported by Bobko [1983] agrees exactly with the formula derived by Bobko and Rieck [1980] for the square of the standard error of the correlation corrected for [Case 2 [Case A]] restriction of range only [Equation 4.35] when  $\rho_{\text{tty}}$  in Equation 4.37 equals one [i.e. when no correction for criterion unreliability occurs].

An analysis of Equation 4.37 by Bobko [1983] results in the following conclusions:

- the standard error associated with the doubly corrected correlation coefficient is larger than the standard error of the uncorrected coefficient;
- the standard error will decrease in direct proportion to the square root of the size of the sample;
- the standard error will increase as  $\rho_{\text{tty}}$  decreases; and
- the standard error will increase as the selection ratio decreases.

The reaction of  $\sigma[\hat{\rho}_c]$  to changes in  $\rho[x, y]$ ,  $K$ ,  $\rho_{\text{tty}}$  and  $n$  is graphically portrayed in Figures 4.51 - 4.56. Figures 4.51 - 4.56 confirm the analytically derived conclusions of Bobko [1983] in as far as they show that  $\sigma[\hat{\rho}_c]$ :

- increases curvilinearly as  $\rho_{\text{tty}}$  decreases;
- increases as  $K$  increases [i.e. the selection ratio decreases];
- decreases as  $n$  increases; and in addition suggest that  $\sigma[\hat{\rho}_c]$
- increases as  $\rho[x, y]$  decreases.

The standard error of the double corrected correlation coefficient peaks at low values of  $\rho_{\text{tty}}$  and  $\rho[x, y]$ .

The reaction of the standard error ratio  $V = \sigma[\hat{\rho}_c] / \sigma[\hat{\rho}]$  to changes in  $\rho[x, y]$ ,  $K$ ,  $\rho_{\text{tty}}$  and  $n$  is graphically portrayed in Figures 4.57 - 4.60. The reaction pattern of  $V$  to changes in the relevant parameters seem to closely correspond to the reaction pattern of  $\sigma[\hat{\rho}_c]$  under similar conditions. Figures 4.57 - 4.60, furthermore, suggest that Bobko [1983] is in error with his unqualified, analytically derived, conclusion that the standard error associated with the double corrected correlation coefficient is larger than the standard error of the uncorrected coefficient.  $V$  in fact drops below unity in a relatively small region of the  $\rho_{\text{tty}}$ ,  $\rho[x, y]$  plane characterised by high values on both the  $\rho_{\text{tty}}$  and  $\rho[x, y]$  axes.

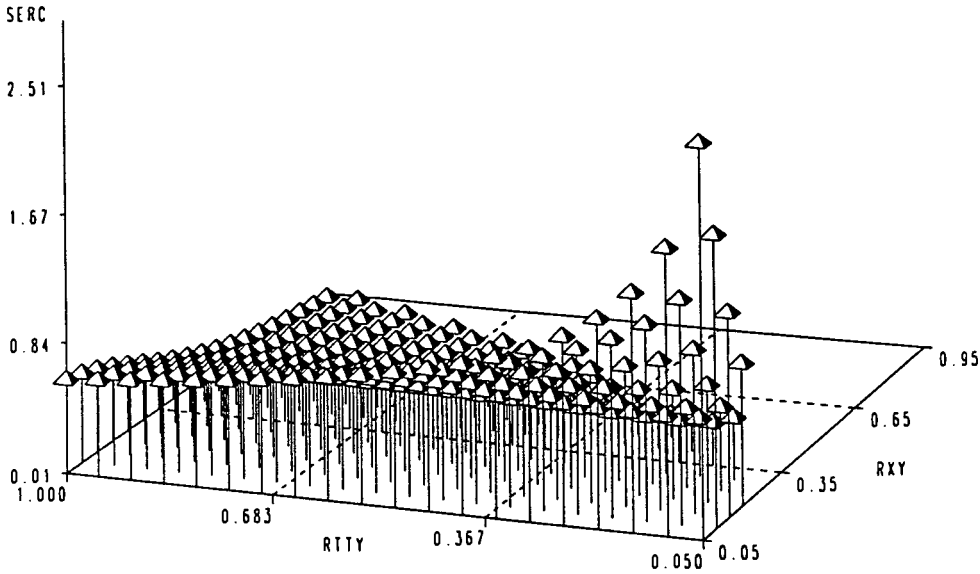


Figure 4.51: The standard error of the validity coefficient jointly corrected for Case 2 [Case A] restriction of range and criterion unreliability as a function of  $\rho[x,y]$  and  $\rho_{ttY}$  for  $n$  fixed at 10 and  $K$  fixed at 2

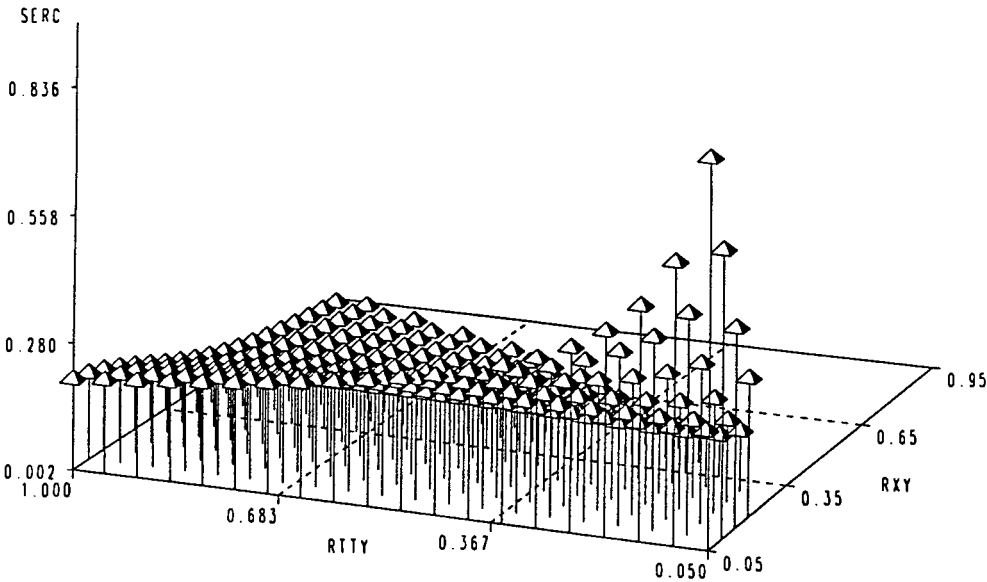


Figure 4.52: The standard error of the validity coefficient jointly corrected for Case 2 [Case A] restriction of range and criterion unreliability as a function of  $\rho[x,y]$  and  $\rho_{ttY}$  for  $n$  fixed at 90 and  $K$  fixed at 2



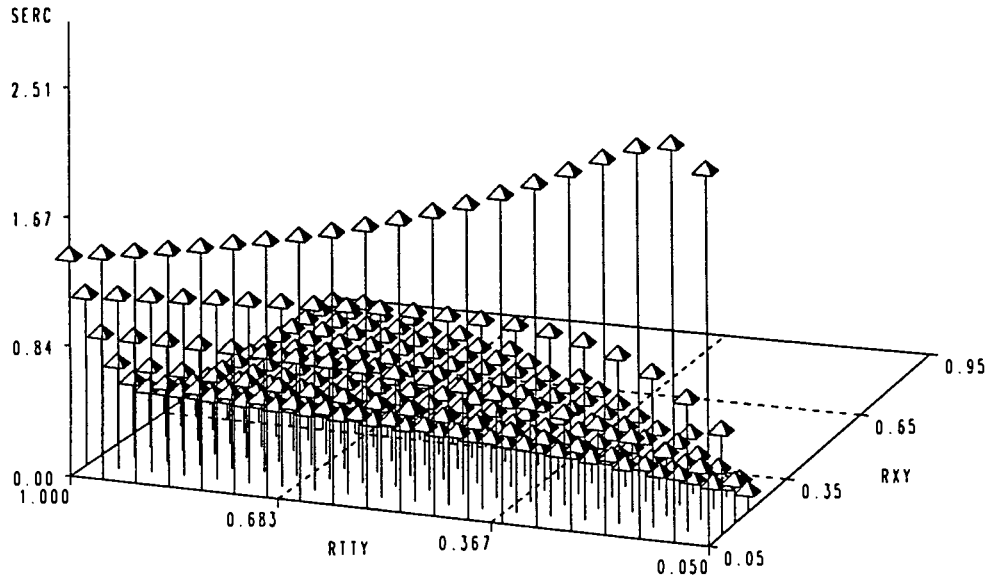


Figure 4.53: The standard error of the validity coefficient jointly corrected for Case 2 [Case A] restriction of range and criterion unreliability as a function of  $\rho[x,y]$  and  $\rho_{ttY}$  for  $n$  fixed at 10 and  $K$  fixed at 5

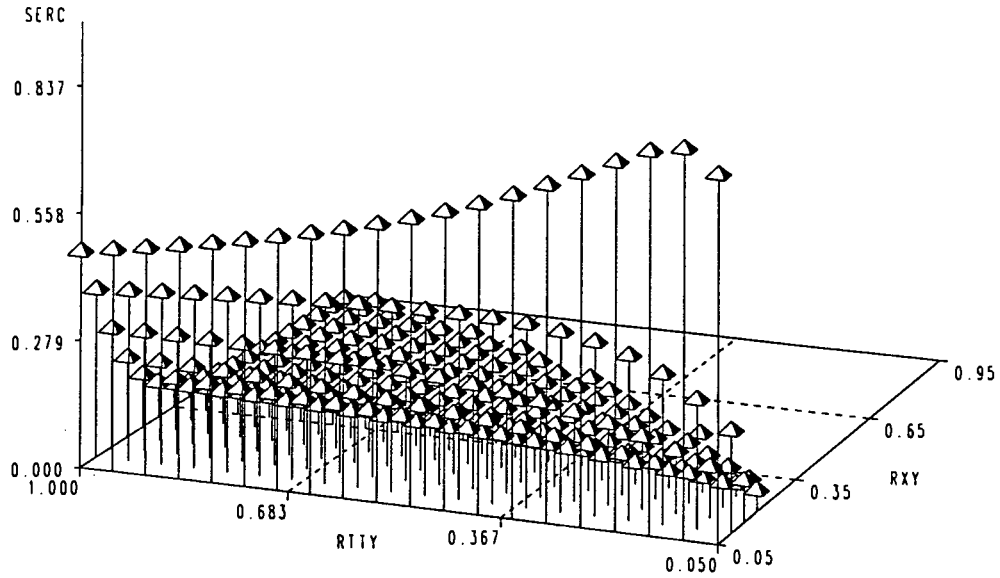


Figure 4.54: The standard error of the validity coefficient jointly corrected for Case 2 [Case A] restriction of range and criterion unreliability as a function of  $\rho[x,y]$  and  $\rho_{ttY}$  for  $n$  fixed at 90 and  $K$  fixed at 5

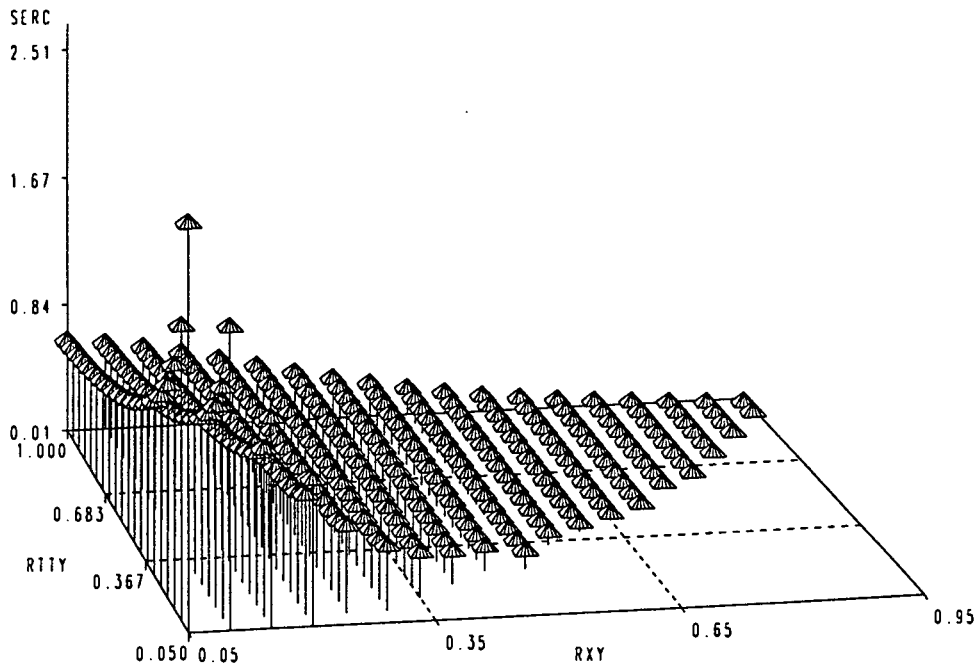


Figure 4.55: The standard error of the validity coefficient jointly corrected for Case 2 [Case A] restriction of range and criterion unreliability as a function of  $\rho[x,y]$  and  $\rho_{ttY}$  for  $n$  fixed at 10 and  $K$  fixed at 2 [rotated through  $60^\circ$ ]

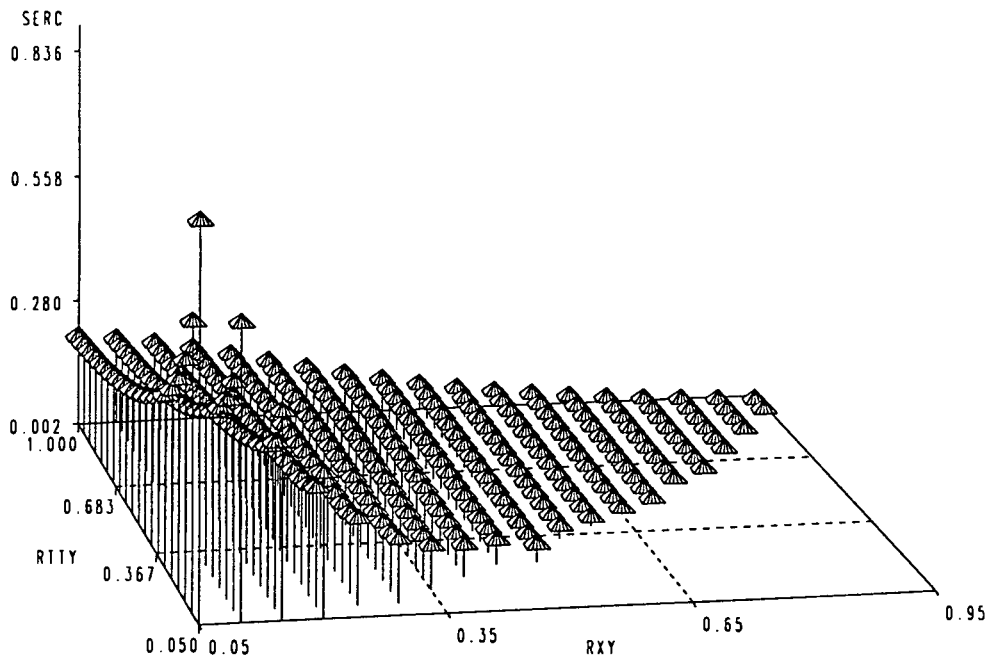


Figure 4.56: The standard error of the validity coefficient jointly corrected for Case 2 [Case A] restriction of range and criterion unreliability as a function of  $\rho[x,y]$  and  $\rho_{ttY}$  for  $n$  fixed at 90 and  $K$  fixed at 2 [rotated through  $60^\circ$ ]

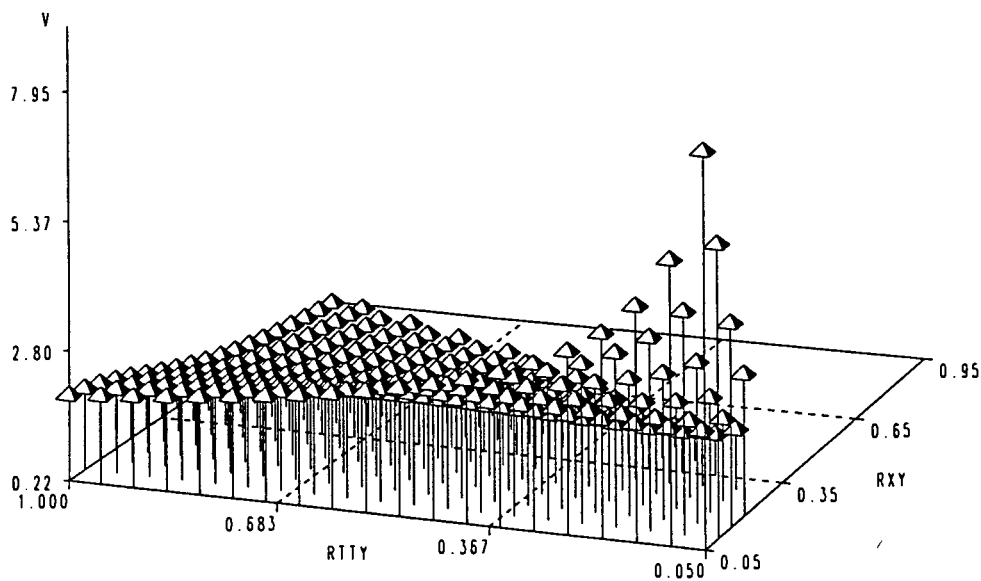


Figure 4.57: The standard error ratio  $V = \sigma[\hat{\rho}_e]/\sigma[\hat{\rho}]$  as a function of  $\rho[x,y]$  and  $\rho_{ttY}$  for  $n$  fixed at 10 and  $K$  fixed at 2

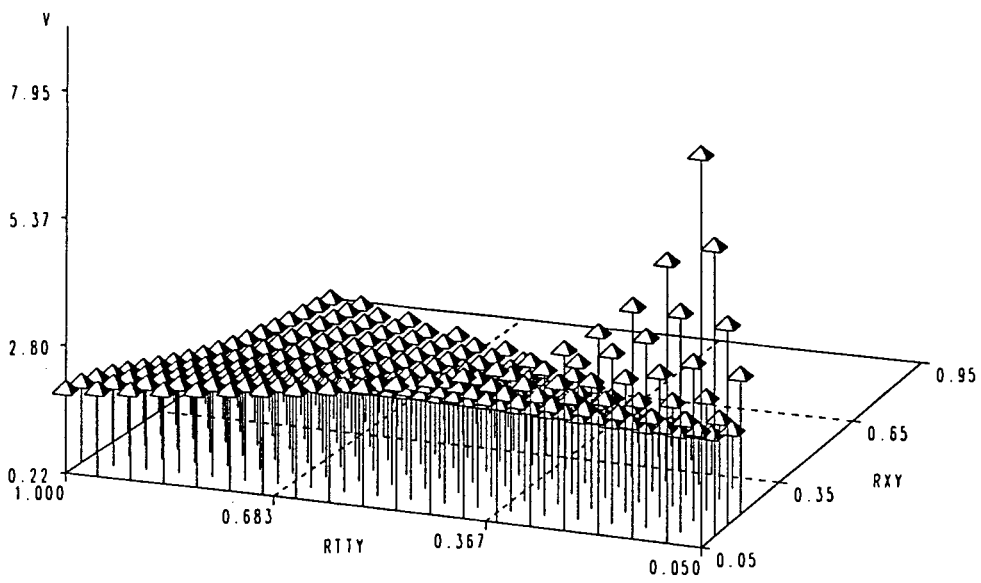


Figure 4.58: The standard error ratio  $V = \sigma[\hat{\rho}_e]/\sigma[\hat{\rho}]$  as a function of  $\rho[x,y]$  and  $\rho_{ttY}$  for  $n$  fixed at 90 and  $K$  fixed at 2

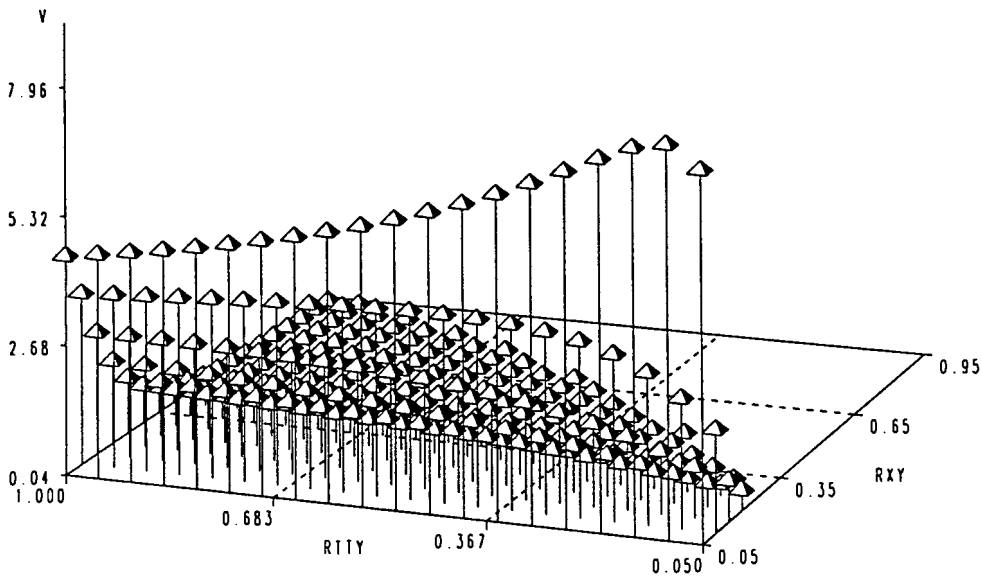


Figure 4.59: The standard error ratio  $V = \sigma[\hat{\rho}_{\cdot}] / \sigma[\hat{\rho}^*]$  as function of  $\rho_{xy}$  and  $\rho_{ttY}$  for  $n$  fixed at 10 and  $K$  fixed at 5

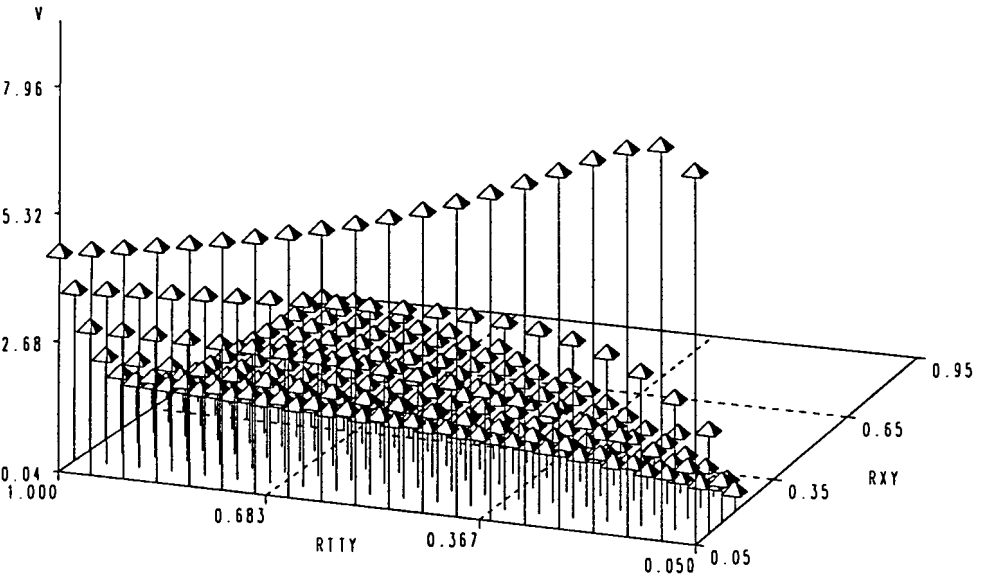


Figure 4.60: The standard error ratio  $V = \sigma[\hat{\rho}_{\cdot}] / \sigma[\hat{\rho}^*]$  as a function of  $\rho_{xy}$  and  $\rho_{ttY}$  for  $n$  fixed at 90 and  $K$  fixed at 5



## CHAPTER 5

### EFFECT OF STATISTICAL CORRECTIONS OF THE VALIDITY COEFFICIENT ON DECISIONS ON THE NULL HYPOTHESES

The purpose of chapter five is to establish the consequences of correcting the Pearson correlation coefficient for attenuation and/or restriction of range on decisions on the “acceptance” or rejection of statistical null hypotheses. In pursuit of this objective, a synopsis of the basic logic underlying [classical/Fisherian] statistical hypothesis/significance testing will first be presented.

#### 5.1 THE LOGIC OF [FISHERIAN] STATISTICAL SIGNIFICANCE TESTING

Let  $\Pi$  represent a finite and bivariate normal population for which a parameter  $\theta$  exists with an unknown value. Let  $\Psi$  represent a random sample of size  $n$  drawn from  $\Pi$  for which the statistic  $\theta^{\wedge}$  is calculated as an unbiased estimate [i.e.  $E[\theta^{\wedge}] = \theta$ ] of the unknown value of the parameter  $\theta$ . Let  $\theta^{\wedge}_b$  represent the obtained sample value of  $\theta^{\wedge}$ . The obtained sample value  $\theta^{\wedge}_b$  is of interest in as far as it provides an estimate of the unknown parameter value. The sample estimate  $\theta^{\wedge}_b$  will, however, still only reflect with less than perfect accuracy, the parameter of interest  $\theta$  due to sampling error. Results  $[\theta^{\wedge}]$  from a random sample  $\Psi$  will only approximate the characteristics  $[\theta]$  of the population  $\Pi$ ,  $\Psi$  was drawn from. Inherent to generalising the sample estimate beyond the confines of the sample is a certain degree of uncertainty. Due to the inherent inability of any sample [ $n < N$ ;  $\sigma^2 > 0$ ] to perfectly represent its parent population [i.e. sampling error] the possibility always exists that the sample statistic is attributable purely to chance [i.e. the possibility always exists that  $\theta = \theta_0$  even though  $\theta^{\wedge} \neq \theta_0$ ]. The question specifically arises whether the sample result  $\theta^{\wedge} = \theta^{\wedge}_b$  permits the inference that  $\theta \neq \theta_0$ ;  $\theta_0 = 0$  but approximately equal to  $\theta^{\wedge}_b$ . The question is thus whether  $\theta^{\wedge} = \theta^{\wedge}_b$  may be considered statistically significant or not.

Whether the obtained value of the statistic may be considered statistically significant depends on the probability  $[\alpha_b]$  of obtaining the sample result, or something more extreme, [i.e. more unlikely] if the hypothesised condition would be true. Statistical significance testing thus investigates the null hypothesis  $H_0: \theta = \theta_0$  by estimating the conditional probability  $\alpha_b = P[\theta^{\wedge} \geq \theta^{\wedge}_b \mid \theta = \theta_0]$  [assuming  $\theta_b > \theta_0$ ]. Should the probability of the obtained sample result conditional on  $H_0$  be sufficiently small, the conjecture contained in  $H_0$  can, albeit with an element of risk, be rejected due to its incompatibility with the empirically derived [i.e. factual] sample result. This, however begs the question when the conditional probability  $\alpha_b$  may be interpreted as small.



The term small [probability] is operationalised in terms of a critical probability [ $\alpha_c$ ]. The magnitude of  $\alpha_c$  is dependent on the combined effect of:

- the chosen significance level [ $\alpha$ ] or probability of [mistakenly] rejecting  $H_0$  when in fact  $H_0$  is true [i.e. the chosen probability level for a Type I error];
- the directional or non-directional nature of  $H_a$ .

Convention dictates  $\alpha = 0,05$  or  $\alpha = 0,01$  as two acceptable levels of risk for a Type I error in deciding on the adequacy of  $H_0$ . Should  $H_0$  be paired with a non-directional alternative hypothesis [i.e.  $H_a: \theta \neq \theta_0$ ],  $\alpha_c = \alpha/2$ . Should  $H_0$ , however, be paired with a directional alternative hypothesis [i.e.  $H_a: \theta > \theta_0$  or  $H_a: \theta < \theta_0$ ],  $\alpha_c = \alpha$ .

Thus the following basic decision rule for classical [Fisherian] hypothesis testing can be stated [assuming  $\theta_b > \theta_0$ ]:

- reject  $H_0$  if  $P[\theta^{\wedge} \geq \theta^{\wedge}_b | \theta = \theta_0] = \alpha_b \leq \alpha_c$ ; and
- fail to reject  $H_0$  if  $P[\theta^{\wedge} \geq \theta^{\wedge}_b | \theta = \theta_0] = \alpha_b > \alpha_c$

The null hypothesis will thus be rejected if the probability of observing the sample result  $\theta^{\wedge} = \theta^{\wedge}_b$  in a sample of size  $n$ , conditional on the assumption that  $\theta = \theta_0$  in the population from which the sample was randomly drawn, is sufficiently small. Conversely, the null hypothesis will/can not be rejected if the probability of observing the sample result  $\theta^{\wedge} = \theta^{\wedge}_b$  in a sample of size  $n$ , conditional on the assumption that  $\theta = \theta_0$  in the population from which the sample was randomly drawn, is still relatively large. Should  $\theta_0$  equal 0, rejection of  $H_0$  would imply the obtained sample value  $\theta^{\wedge} = \theta^{\wedge}_b$  to be statistically significant [ $p < \alpha_c$ ]

Given two possible decisions and two possible true states of nature four possible outcomes for any hypothesis test are therefore implied. These four possible outcomes and their associated probabilities are shown in Table 5.1.

Table 5.1: Possible outcomes of hypothesis testing and their associated probabilities

TRUE STATE OF NATURE		
DECISION	$H_0$ TRUE	$H_a$ TRUE
REJECT $H_0$	TYPE I ERROR WITH PROBABILITY= $\alpha$	CORRECT DECISION WITH PROBABILITY= $1-\beta$
"ACCEPT" $H_0$	CORRECT DECISION WITH PROBABILITY= $1-\alpha$	TYPE II ERROR WITH PROBABILITY= $\beta$

Figure 5.1 provides a graphical portrayal of the probabilities associated with the four possible hypothesis testing outcomes for  $H_0: \theta = \theta_0$  against  $H_a: \theta > \theta_0$  [which logically should be the typical

situation in validation research] when  $\theta$  is in fact equal to  $\theta_t$  [ $\theta_t > \theta_0$ ]. Let  $\theta^*_c$  represent a critical score such that  $P[\theta^* \geq \theta^*_c | \theta = \theta_0] = \alpha_c = \alpha$ . Let  $\theta^*_b$  still represent the observed value of the statistic  $\theta^*$  and let  $\alpha_b$  represent the probability of observing  $\theta^*_b$  in a random sample drawn from a population with  $\theta$  equal to  $\theta_0$  [i.e. conditional on  $H_0$ ].

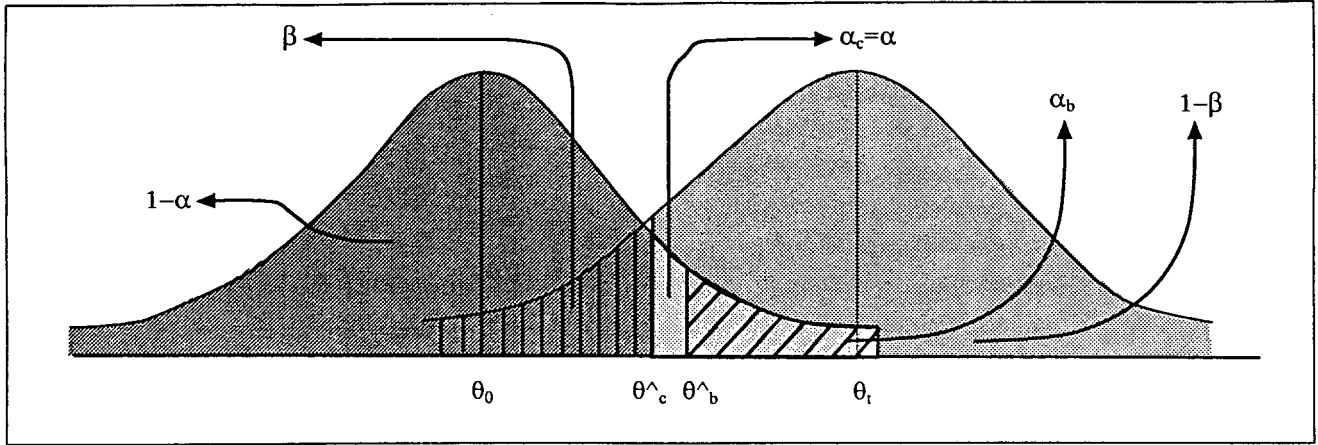


Figure 5.1: Probabilities associated with the possible hypothesis test outcomes [directional alternative hypothesis]

## 5.2 EFFECT OF CORRECTIONS FOR ATTENUATION AND/OR RESTRICTION OF RANGE TO THE PEARSON CORRELATION COEFFICIENT ON DECISIONS ON STATISTICAL NULL HYPOTHESES

The question could be phrased whether corrections to the obtained validity coefficient should be implemented or not. Stated as such it remains a scientifically unanswerable, metaphysical and generally problematic question. The problem firstly lies with the term "should", which implies that, in the absence of an objective criterion, some value judgement would be required to answer the question. Should the term be operationalised in terms of one or more measurable criteria, this problem disappears. A more fundamental problem remains, however, namely whether the question posed is logical/rational to start with. The question as posed implies a choice between the uncorrected validity coefficient  $\rho^*$  and the corrected validity coefficient  $\rho^*_c$ . A choice between  $\rho^*$  and  $\rho^*_c$ , however, makes no sense since both constitute meaningful but qualitatively different, although related, quantities. The one thus cannot serve as a logical substitute for the other. However, if the question is rephrased by inquiring into the individual and combined effect of corrections for attenuation and restriction of range on decisions on the validity of performance hypotheses, a more rational problem seems to emerge.

The preceding discussion provides the necessary conceptual framework needed to examine the possible effects of corrections for attenuation and/or restriction of range to the Pearson correlation coefficient on decisions on the truth of statistical hypotheses. Two quantities that were referred to in the foregoing discussion seem to have particular relevance for this [rephrased] question, namely:

- the empirically derived exceedence probabilities [or achieved significance level, ASL]  
 $\alpha_b = P[\theta^{\wedge} \geq \theta^{\wedge}_b \mid \theta = \theta_0]$  for  $\theta = \rho$  and  $\theta = \rho_c$ ;
- the theoretical probabilities  $[1-\beta] = P[\theta^{\wedge} \geq \theta^{\wedge}_c \mid \theta = \theta_1]$  for  $\theta = \rho$  and  $\theta = \rho_c$  or power of the tests of  $H_0$ .

Traditional Fisherian decision rule dictates that the null hypothesis may be rejected if  $\alpha_b \leq \alpha_c$ . Assume for the sake of the immediate argument that on the basis of the pre-correction analysis,  $H_0$  could not be rejected. Clearly, should the value specified under  $H_0$ , the significance level and the directional/non-directional nature of  $H_a$  [and consequently  $\alpha_c$ ] be kept constant, a decrease in  $\alpha_b$  would improve the chances of rejecting  $H_0$ . Under conditions where both  $\theta^{\wedge}_b$  and  $\sigma^2[\theta^{\wedge}]$  increase,  $\alpha_b$  would remain unaltered only if the change/increase in the correlation coefficient expressed in terms of the change/increase in the standard error [i.e.  $\Delta\theta^{\wedge}/\Delta\sigma^2[\theta^{\wedge}]$ ] equals the initial ratio  $\theta^{\wedge}/\sigma^2[\theta^{\wedge}]$ . By the same logic a decrease in  $\alpha_b$  would only be possible if the ratio  $\Delta\theta^{\wedge}/\Delta\sigma^2[\theta^{\wedge}]$  exceeds the initial ratio  $\theta^{\wedge}/\sigma^2[\theta^{\wedge}]$ . Stated differently, a decrease in  $\alpha_b$  would only be possible if the subsequent [post-correction] ratio  $\theta^{\wedge}_c/\sigma^2[\theta^{\wedge}_c]$  exceeds the initial ratio  $\theta^{\wedge}/\sigma^2[\theta^{\wedge}]$ .

Applying the preceding argument to the [rephrased] question stated earlier thus leads to the conclusion that one possible effect of correcting the Pearson correlation coefficient for attenuation and/or restriction of range could be that the probability of observing the uncorrected correlation  $\rho^{\wedge}$  in a sample of size  $n$  drawn from a population where the two correlated variables are independent is either smaller or larger than the conditional probability of observing the corrected correlation under the same conditions. The effect of the corrections could thus be that of either increasing, decreasing or leaving unaltered the a posteriori probability of rejecting  $H_0$ . The pivotal question thus becomes whether the ratio of the change in the value of the correlation to the change in the standard error of the correlation [i.e.  $(\rho^{\wedge}_c - \rho^{\wedge})/(\sigma[\rho^{\wedge}_c] - \sigma[\rho^{\wedge}])$ ] is bigger or smaller than the ratio  $\rho^{\wedge}/\sigma[\rho^{\wedge}]$ . Assuming  $Z = \rho^{\wedge}/\sigma[\rho^{\wedge}]$  to follow a standard normal  $[0, \sigma[\rho^{\wedge}]]$  distribution under  $H_0: \rho = 0$  [Kendall & Stuart, 1977] and assuming  $Z_c = \rho^{\wedge}_c/\sigma[\rho^{\wedge}_c]$  to, likewise, follow a standard normal  $[0, \sigma[\rho^{\wedge}_c]]$  distribution under  $H_0: \rho_c = 0$  [Bobko & Rieck, 1980; Moran, 1970], it thus follows that the decisive factor is the magnitude of  $G = Z_c/Z$ . If  $G$  equals unity,  $\alpha_b$  remains unaffected. However, if  $G > 1$  then  $\alpha_b$  decreases and conversely, if  $G < 1$  then  $\alpha_b$  increases [assuming both  $\rho^{\wedge}$  and  $\rho^{\wedge}_c$  to be positive].

Previously [see chapter 1] it was argued that corrections to the validity coefficient would be important if they change decisions on the validity of hypotheses explaining variance in performance at a cost substantially lower than the value of the altered decision. In operational terms the critical question

consequently is [assuming  $G \neq 1$  and assuming correction cost to be negligible] under which conditions the change in  $\alpha_b$  [both increases and decreases] produces movement past the critical value  $\alpha_c$ . The relevant parameters that need to be considered when defining the aforementioned conditions would be those contained in the expression for the standard error of the corrected correlation coefficient and those affecting  $\alpha_c$ .

The behaviour of  $G$  under different values of the relevant parameters could be examined through computer generated plots of  $G$  against those parameters. Such a plot would indicate whether corrections to the correlation coefficient affect  $\alpha_b$ , how they affect  $\alpha_b$  and under what conditions. Such a plot would, however, be incapable of indicating whether the change in standard score [or  $\alpha_b$ ] is sufficient to affect the decision on  $H_0$ . This shortcoming could, however, be circumvented by plotting  $G$  on a continuous and a discrete/nominal scale simultaneously through the utilisation of appropriate character or colour codes. Table 5.2 indicates the nominal scale [ $G_T$ ] used to examine the effect of corrections to the validity coefficient on  $\alpha_b$ .

Table 5.2: The nominal scale used to scale  $G$  to reflect effect of changes in  $Z_b$  on the decision on  $H_0$

DECISION $H_{02}: \rho_c=0$		
DECISION $H_{01}: \rho=0$	REJECT $H_0$	"ACCEPT $H_0$ "
REJECT $H_0$	SAME DECISION $G_T = 1$ [CUBE]	CHANGED DECISION $G_T = 2$ [HEART]
"ACCEPT" $H_0$	CHANGED DECISION $G_T = 3$ [SPADE]	SAME DECISION $G_T = 4$ [FLAG]

A change in decision brought about by correcting the correlation coefficient does not, however, necessarily represent an advantage. Whether a change does represent an advantage depends on the initial decision vis-a-vis the true [but unknown] state of nature. Table 5.1, therefore needs to be superimposed on Table 5.2. The lower left quadrant [i.e.  $G_T = 3$ ], rather than one of the remaining three quadrants, represents the beneficial decision related outcome resulting from correcting the correlation coefficient for criterion unreliability and/or restriction of range if  $H_0$  is in fact false. Conversely, the upper right quadrant [i.e.  $G_T = 2$ ] represents the beneficial decision related outcome resulting from correcting the correlation coefficient for criterion unreliability and/or restriction of range if  $H_0$  is in fact true. Should the cost of implementing the correction approach zero, outcomes  $G_T = 1$  and  $G_T = 4$  would also represent acceptable outcomes under the aforementioned two true states of nature.



The concept of statistical power refers to the apriori probability [that is the pre-analysis/pre-decision probability] of rejecting the null hypothesis when  $H_0$  is in fact false [Cohen, 1977; Lipsey, 1990; Toothaker, 1986]. With reference to Figure 5.1, power refers to  $[1-\beta] = P[\theta^{\wedge} \geq \theta^{\wedge}_c | \theta = \theta_t]$  where  $\theta^{\wedge}_c$  represent a critical score such that  $P[\theta^{\wedge} \geq \theta^{\wedge}_c | \theta = \theta_0] = \alpha_c$  and  $\theta_t$  the value for the unknown parameter  $\theta$  assumed under  $H_a$ . The power of a statistical hypothesis test depends on the following four parameters [Cohen, 1977; Toothaker, 1986]:

- the chosen significance level  $[\alpha]$ ;
- the directional or non-directional nature of  $H_a$ ;
- the sample size  $[n]$ ; and
- the actual value of the population parameter  $\theta$  or effect size [ES].

Assuming  $K = \sigma[X]/\sigma[x]$  always greater than unity, it follows that the correlation coefficient corrected for either criterion unreliability or restriction of range or both, will always exceed the uncorrected correlation coefficient. Furthermore, as was indicated in chapter 4, the standard error of the corrected correlation coefficient mostly exceeds the standard error of the uncorrected correlation coefficient. The effect of statistical corrections on the power of a statistical hypothesis test on the corrected correlation coefficient will thus depend on how the specification of effect size [ES] is approached. If the hypothesised effect of the independent variable on the dependent variable is translated into an appropriate statistical index [Cohen, 1977] independent of criterion reliability and/or restriction of range considerations, power of the hypothesis test on the corrected coefficient will be lower.

Let  $\rho_i$  denote the value of the validity coefficient assumed under  $H_a$  [i.e. the effect size].

Let  $Z_\alpha$  denote the standardised ordinate corresponding to the critical uncorrected correlation  $\rho^{\wedge}_{crit}$  and corrected correlation  $\rho^{\wedge}_{ccrit}$  under  $H_0$  such that  $P[\rho^{\wedge} \geq \rho^{\wedge}_{crit} | \rho = 0] = \alpha_c$  and  $P[\rho^{\wedge}_c \geq \rho^{\wedge}_{ccrit} | \rho_c = 0] = \alpha_c$

Consequently:

$$\rho^{\wedge}_{crit} = Z_\alpha \sigma[\rho^{\wedge}] \text{ and } \rho^{\wedge}_{ccrit} = Z_\alpha \sigma[\rho^{\wedge}_c]$$

Let  $Z_{cc}$  denote the critical cutoff score  $[\rho^{\wedge}_{ccrit}]$  under  $H_0$  [for a given  $\alpha_c$ ] expressed as a standard score in the sampling distribution under  $H_a$  for the corrected correlation

Therefore:

$$\begin{aligned} Z_{cc} &= (\rho^{\wedge}_{ccrit} - \rho_i) / \sigma[\rho^{\wedge}_c] \\ &= (Z_\alpha \sigma[\rho^{\wedge}_c] - \rho_i) / \sigma[\rho^{\wedge}_c] \end{aligned}$$

$$= Z_{\alpha} - (\rho_t / \sigma[\hat{\rho}_c])$$

Let  $Z_c$  denote the critical cutoff score  $[\hat{\rho}_{crit}]$  under  $H_0$  [for a given  $\alpha_c$ ] expressed as a standard score in the sampling distribution under  $H_a$  for the uncorrected correlation

Therefore:

$$\begin{aligned} Z_c &= (\hat{\rho}_{crit} - \rho_t) / \sigma[\hat{\rho}] \\ &= (Z_{\alpha} \sigma[\hat{\rho}] - \rho_t) / \sigma[\hat{\rho}] \\ &= Z_{\alpha} - (\rho_t / \sigma[\hat{\rho}]) \end{aligned}$$

The lower power is therefore attributable to the fact that the critical cutoff score  $[\hat{\rho}_{ccrit}]$  under  $H_0$  [for a given  $\alpha_c$ ] expressed as a standard score in the sampling distribution under  $H_a$  for the corrected correlation  $[(Z_{\alpha} \sigma[\hat{\rho}_c] - \rho_t) / \sigma[\hat{\rho}_c] = Z_{\alpha} - (\rho_t / \sigma[\hat{\rho}_c])]$  will be bigger [i.e. less negative; assuming  $\rho_t > 0$ ] than the critical cutoff score under  $H_0$  [for the same  $\alpha_c$ ] expressed as a standard score in the sampling distribution under  $H_a$  for the uncorrected correlation  $[(Z_{\alpha} \sigma[\hat{\rho}] - \rho_t) / \sigma[\hat{\rho}] = Z_{\alpha} - (\rho_t / \sigma[\hat{\rho}])]$ , since  $Z_{\alpha}$  and  $\rho_t$  remain constant but  $\sigma[\hat{\rho}_c] > \sigma[\hat{\rho}]$ . This argument is portrayed graphically in Figure 5.2.

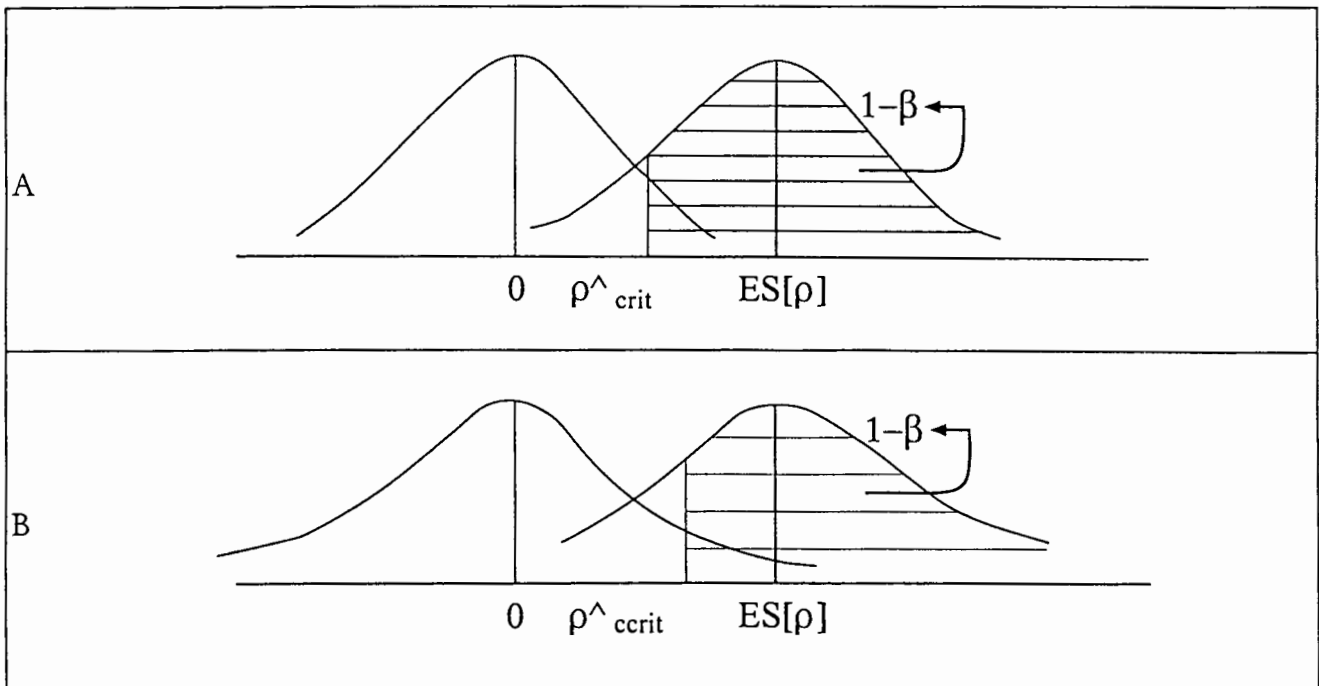


Figure 5.2: Effect of corrections to the correlation coefficient on statistical power when assuming a constant effect size [graph A represents the situation for the uncorrected correlation and graph B for the corrected correlation]



The same conclusion would result if ES is specified once for the unselected applicant population and a perfectly reliable criterion only. Cohen [1977, pp. 9-10] interprets the concept effect size as follows:

Without intending any necessary implications of causality, it is convenient to use the phrase 'effect size' to mean 'the *degree* to which the phenomenon is present in the population,' or 'the degree to which the null hypothesis is false.' ... the ES can itself be treated as a parameter which takes the value zero when the null hypothesis is true and some other specific nonzero value when the null hypothesis is false, and in this way the ES serves as an index of degree of departure from the null hypothesis.

The foregoing definition suggests that the hypothesised effect of the independent variable on the dependent variable cannot be translated into an appropriate statistical index [Cohen, 1977] independent of criterion reliability and/or restriction of range considerations, since ES is contingent on a definition of the population and the attribute which is hypothesised to be affected by one or more effects. This implies that an unambiguous outline of the applicant population to which the correlation coefficient is meant to generalise as well as the affected attribute, is required. This in turn would imply that the specific value of  $\rho$  postulated under  $H_a$  for the selected applicant population and/or applicant population for whom the criterion measurements contains no random error [i.e.  $\rho_t$ ] will have to differ from the specific value for  $\rho$  assumed under  $H_a$  for the unrestricted applicant population and/or applicant population for whom the criterion measurements does contain random error. Thus the specific value of  $\rho$  postulated under  $H_a$  for the selected applicant population and/or applicant population for whom the criterion measurements does contain random error will have to be corrected to obtain the estimate of  $\rho_c$  that should be assumed under  $H_a$ .

The fundamental underlying question, however, still is on the magnitude of the critical cutoff score [ $\hat{\rho}_{crit}$ ] under  $H_0$  [for a given  $\alpha_c$ ] expressed as a standard score in the sampling distribution under  $H_a$  for the corrected correlation [ $(Z_\alpha \sigma[\hat{\rho}_c] - \rho_c) / \sigma[\hat{\rho}_c] = Z_\alpha - (\rho_c / \sigma[\hat{\rho}_c])$ ] relative to the critical cutoff score [ $\hat{\rho}_{crit}$ ] under  $H_0$  [for the same  $\alpha_c$ ] expressed as a standard score in the sampling distribution under  $H_a$  for the uncorrected correlation [ $(Z_\alpha \sigma[\hat{\rho}] - \rho_t) / \sigma[\hat{\rho}] = Z_\alpha - (\rho_t / \sigma[\hat{\rho}])$ ]. If corrections applied to the correlation coefficient would have the effect of increasing the magnitude of the critical cutoff score under  $H_0$  expressed as a standard score in the sampling distribution under  $H_a$ , the power of the hypothesis test would thereby necessarily be diminished. Since  $Z_\alpha$  is a constant, the aforementioned question thus translates to the question how the ratio  $\rho_t / \sigma[\hat{\rho}]$  compares to the ratio  $\rho_c / \sigma[\hat{\rho}_c]$ . The critical ratio is thus  $(\rho_t / \sigma[\hat{\rho}]) / (\rho_c / \sigma[\hat{\rho}_c])$ . Should  $(\rho_t / \sigma[\hat{\rho}]) / (\rho_c / \sigma[\hat{\rho}_c])$  equal unity, the power of the test of the hypothesis  $H_0: \rho = 0$  would equal the power of the test of the hypothesis  $H_0: \rho_c = 0$ . However, should the ratio  $(\rho_t / \sigma[\hat{\rho}]) / (\rho_c / \sigma[\hat{\rho}_c]) < 1$ , power would be positively affected since it would imply that the critical cutoff under  $H_0$  would translate to a more extreme standard score under  $H_a$  for the corrected correlation coefficient than for the uncorrected coefficient. Furthermore, since both parameters constituting the ratio vary across ratios, the

relationship between the two ratios would depend on the change in  $\rho$  relative to the change in the standard error [i.e.  $\Delta\hat{\rho}/\Delta\sigma[\hat{\rho}]$ ] vis-a-vis the initial ratio  $\hat{\rho}/\sigma[\hat{\rho}]$ . Should  $\Delta\hat{\rho}/\Delta\sigma[\hat{\rho}]$  equal  $\hat{\rho}/\sigma[\hat{\rho}]$  power would be unaffected. This argument is portrayed graphically in Figure 5.3.

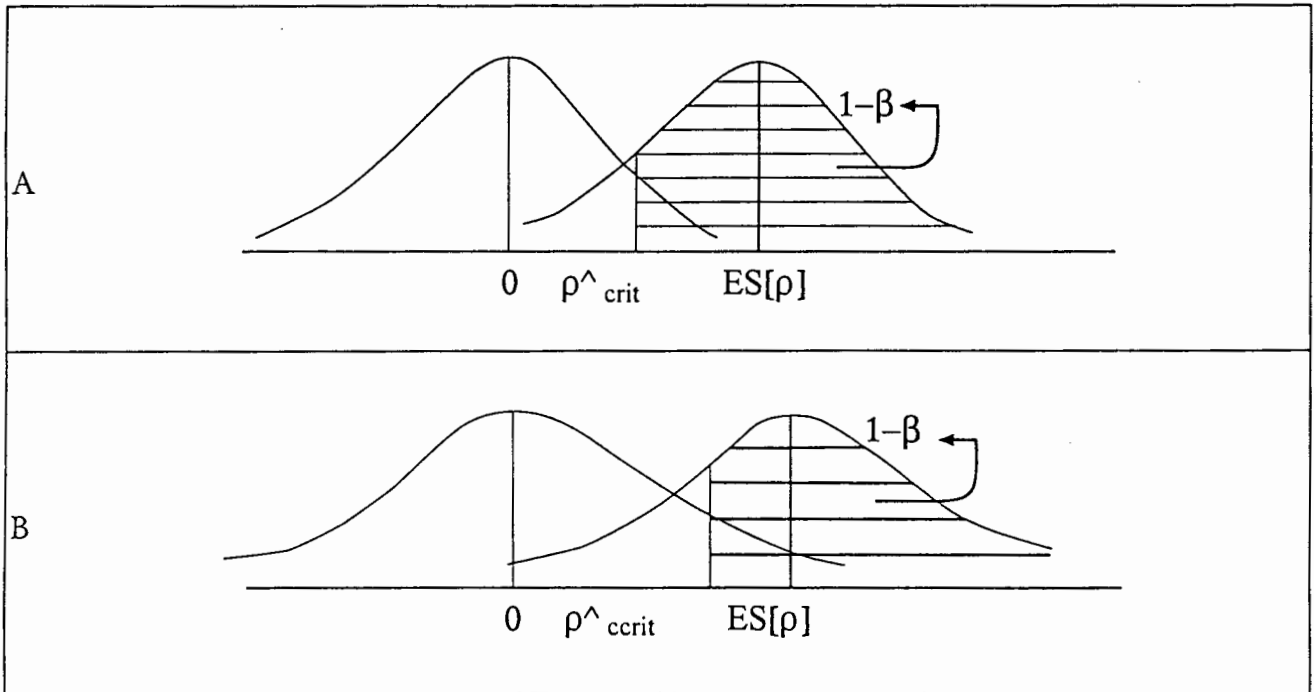


Figure 5.3: Effect of corrections to the correlation coefficient on statistical power when assuming an adjusted effect size [graph A represents the situation for the uncorrected correlation and graph B for the corrected correlation]

Theoretically the behaviour of the change ratio or the ratio  $(\hat{\rho}_i/\sigma[\hat{\rho}^i])/(\hat{\rho}_{ic}/\sigma[\hat{\rho}^i_c])$  under different relevant conditions could be examined analytically. However, if this should prove to be a too formidable task, the behaviour of  $J = (\hat{\rho}_i/\sigma[\hat{\rho}^i])/(\hat{\rho}_{ic}/\sigma[\hat{\rho}^i_c])$  under different values of the relevant parameters could be examined through computer generated plots of  $J$  against those parameters. Such a plot would indicate whether corrections to the correlation coefficient affect  $(1-\beta)$ , how they affect  $(1-\beta)$  and under what conditions.

### 5.3 EFFECT OF CORRECTIONS FOR ATTENUATION AND/OR RESTRICTION OF RANGE TO THE PEARSON CORRELATION COEFFICIENT ON THE EMPIRICALLY DERIVED EXCEEDENCE PROBABILITIES $\alpha_b$ [OR ACHIEVED SIGNIFICANCE LEVEL, ASL]

The behaviour of  $G = (\hat{\rho}_c/\sigma[\hat{\rho}_c])/(\hat{\rho}/\sigma[\hat{\rho}]) = Z_c/Z$  under different values of the relevant parameters was examined through SAS generated three-dimensional scatter plots of  $G$  against those parameters. Such a plot has the capability of indicating whether corrections to the correlation coefficient affect  $\alpha_b$ , how they affect  $\alpha_b$  and under what conditions.

By plotting  $G$  on a continuous and a discrete/nominal scale  $G_T$  simultaneously through the utilisation of the character codes indicated in Table 5.2, the scatter plots obtained the additional capacity of displaying whether the change in standard score [or  $\alpha_b$ ] is sufficient to affect the decision on  $H_0$ .

#### 5.3.1 Effect Of Corrections For Attenuation To The Pearson Correlation Coefficient On The Empirically Derived Exceedence Probabilities $A_b$

##### 5.3.1.1 The Reliability Coefficient Given A Priori By Theoretical Assumption Or Previously Accepted Knowledge

Figure 5.4 and Figure 5.5 presents a visual description of the behaviour of  $G = (\hat{\rho}_c/\sigma[\hat{\rho}_c])/(\hat{\rho}/\sigma[\hat{\rho}]) = Z_c/Z$  as a function of  $\rho_{ttY}$ , the pre-correction correlation coefficient and sample size  $n$  when the reliability coefficient  $\rho_{ttY}$  is given a priori by theoretical assumption or previously accepted knowledge. The focus is thus on the behaviour of  $G$  when the standard error of the partially disattenuated correlation is calculated via the Bobko and Rieck [1980] Case III expression. Figures 5.4 and 5.5 indicate that  $G = 1$  for all permissible combinations of  $\hat{\rho}[X,Y]$  and  $\rho_{ttY}$ . This would imply that the probability of observing the corrected correlation  $\hat{\rho}[X,Y]$  in a sample of size  $n$  drawn from a population where the two variables being correlated are in fact independent, is equal to the conditional probability of observing the uncorrected correlation under the same conditions. The correction thus leaves the a posteriori probability of rejecting  $H_0$  unaltered at a given sample size. This finding corroborates the analytically derived conclusion of Bobko and Rieck [1980] that the statistical significance tests are asymptotically identical if  $\rho_{ttY}$  is known a priori [See Equation 5.1].

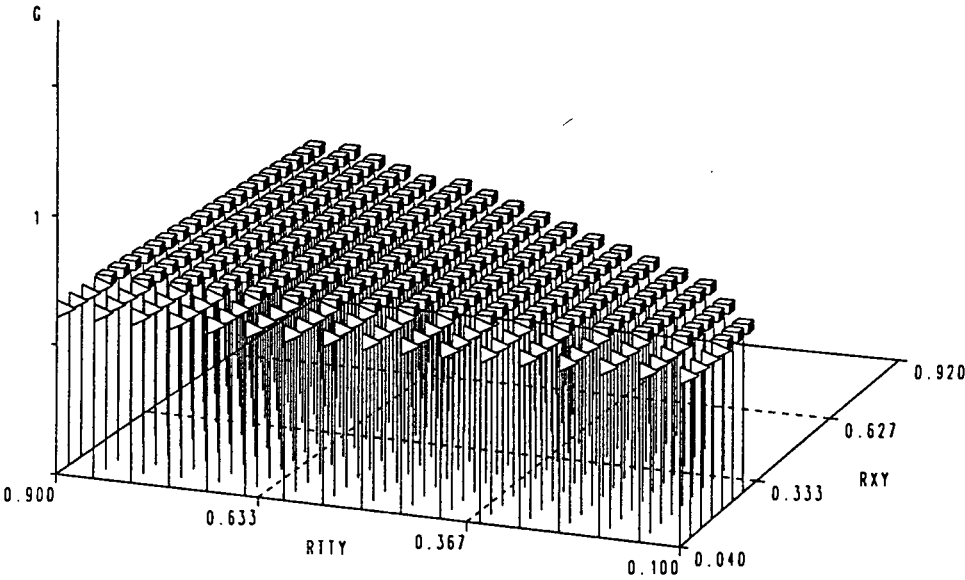


Figure 5.4:  $G = (\hat{\rho}_c / \sigma[\hat{\rho}_c]) / (\hat{\rho} / \sigma[\hat{\rho}]) = Z_c / Z$  as a function of  $\rho_{ttY}$ , the pre-correction correlation coefficient and sample size  $n$  when the reliability coefficient  $\rho_{ttY}$  is given a priori by theoretical assumption or previously accepted knowledge;  $n$  is fixed at 90

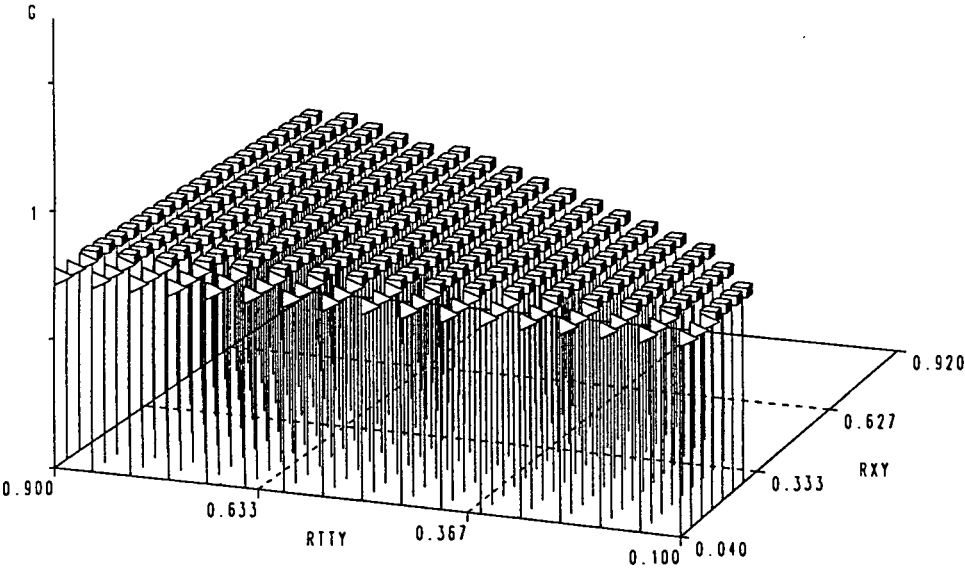


Figure 5.5:  $G = (\hat{\rho}_c / \sigma[\hat{\rho}_c]) / (\hat{\rho} / \sigma[\hat{\rho}]) = Z_c / Z$  as a function of  $\rho_{ttY}$ , the pre-correction correlation coefficient and sample size  $n$  when the reliability coefficient  $\rho_{ttY}$  is given a priori by theoretical assumption or previously accepted knowledge;  $n$  is fixed at 120

$$G = (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}])$$

$$= (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c]) (\sigma[\rho^{\wedge}] / \rho^{\wedge}) \text{-----} 5.1.1$$

Substituting Equation 4.13 and Equation 3.2 in Equation 5.1.1

$$G = ((\rho^{\wedge} / \sqrt{\rho_{ttY}})(1/\sigma[\rho^{\wedge}_c]))((1-\rho^{\wedge 2})/\sqrt{n})(1/\rho^{\wedge}) \text{-----} 5.1.2$$

Canceling out the term  $\rho^{\wedge}$  and substituting Equation 4.25 in Equation 5.1.2:

$$G = (1/\sqrt{\rho_{ttY}})(1/(1/\sqrt{n})(1/\sqrt{\rho_{ttY}})(1-\rho^{\wedge 2}))((1-\rho^{\wedge 2})/\sqrt{n}) \text{-----} 5.1.3$$

Simplifying Equation 5.1.3:

$$G = (1/\sqrt{\rho_{ttY}})(\sqrt{n})(\sqrt{\rho_{ttY}})(1/(1-\rho^{\wedge 2}))((1-\rho^{\wedge 2})/\sqrt{n})$$

$$= 1 \text{-----} 5.1$$

Therefore it follows that  $Z_c = Z$ .

### 5.3.1.2 The Reliability Coefficient Obtained From An Independent Data Set

Figures 5.6 - 5.9 depict the reaction of  $G$  to changes in the initial obtained correlation coefficient, reliability coefficient and sample size when the reliability is obtained from an independent data set. The focus is thus on the behaviour of  $G$  when the standard error of the partially disattenuated correlation is calculated via the Bobko & Rieck [1980] Case II expression. Figures 5.6 - 5.9 indicate that  $G < 1$  for all permissible combinations of  $\rho^{\wedge}[X,Y]$  and  $\rho_{ttY}$  and for all values of  $n$ . This would imply that the probability of observing the corrected correlation  $\rho^{\wedge}_c$  in a sample of size  $n$  drawn from a population where the two variables being correlated are in fact independent, is consistently greater than the conditional probability of observing the uncorrected correlation under the same conditions. The correction thus reduces the a posteriori probability of rejecting  $H_0$  for a given sample size. The increase in  $\alpha_b$  produced by the correction for attenuation has the effect of changing some significant uncorrected correlations into insignificant partially disattenuated correlations. If the three dimensional plots are rotated to provide a view from the  $\rho^{\wedge}[X,Y]$  axis, the region in which the correction changes the significance test outcome can be seen more clearly. Figure 5.10 and Figure 5.11 suggest [see Table 5.2] the change in decision outcome to occur in a specific region on the  $\rho^{\wedge}[X,Y]$  axis, as a function of sample size, for all values of  $\rho_{ttY}$ . As sample size increases the region shifts to lower values of



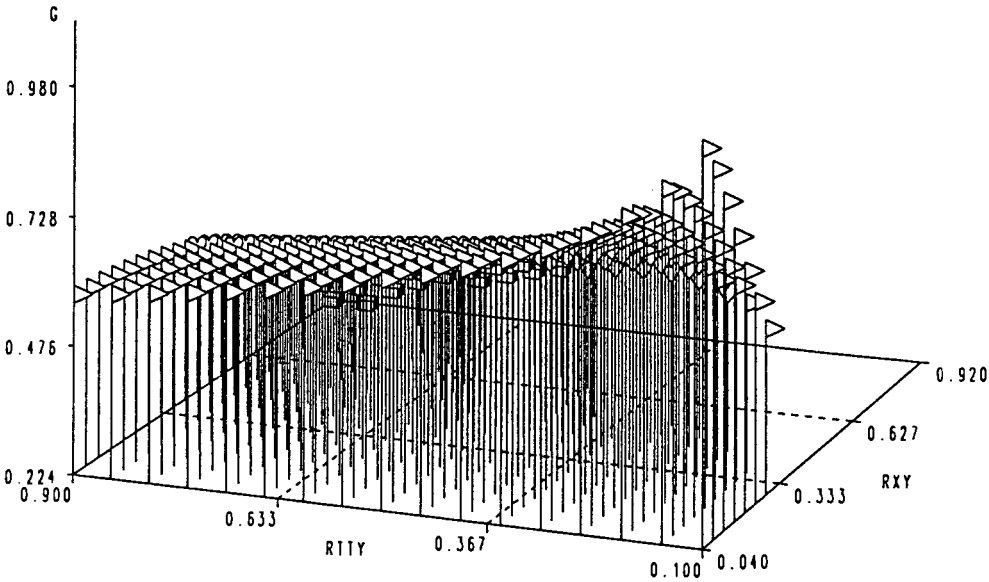


Figure 5.6:  $G = (\hat{\rho}_c/\sigma[\hat{\rho}_c]) / (\hat{\rho}/\sigma[\hat{\rho}]) = Z_c/Z$  as a function of  $\rho_{ttY}$ , the pre-correction correlation coefficient and sample size  $n$  when the reliability coefficient  $\rho_{ttY}$  is obtained from an independent data set;  $n$  is fixed at 10

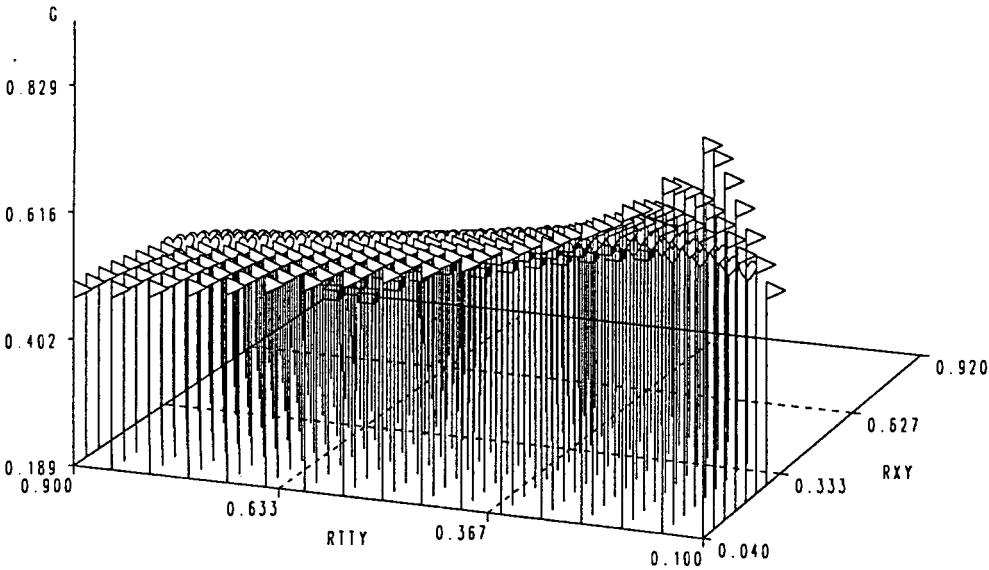


Figure 5.7:  $G = (\hat{\rho}_c/\sigma[\hat{\rho}_c]) / (\hat{\rho}/\sigma[\hat{\rho}]) = Z_c/Z$  as a function of  $\rho_{ttY}$ , the pre-correction correlation coefficient and sample size  $n$  when the reliability coefficient  $\rho_{ttY}$  is obtained from an independent data set;  $n$  is fixed at 20



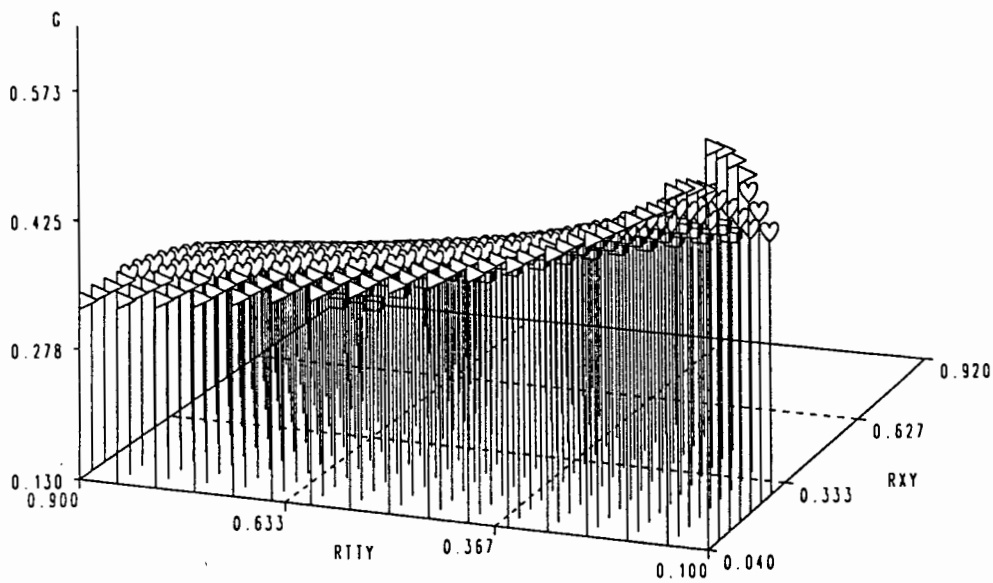


Figure 5.8:  $G = (\hat{\rho}_c / \sigma[\hat{\rho}_c]) / (\hat{\rho} / \sigma[\hat{\rho}]) = Z_c / Z$  as a function of  $\rho_{ttY}$ , the pre-correction correlation coefficient and sample size  $n$  when the reliability coefficient  $\rho_{ttY}$  is obtained from an independent data set;  $n$  is fixed at 90

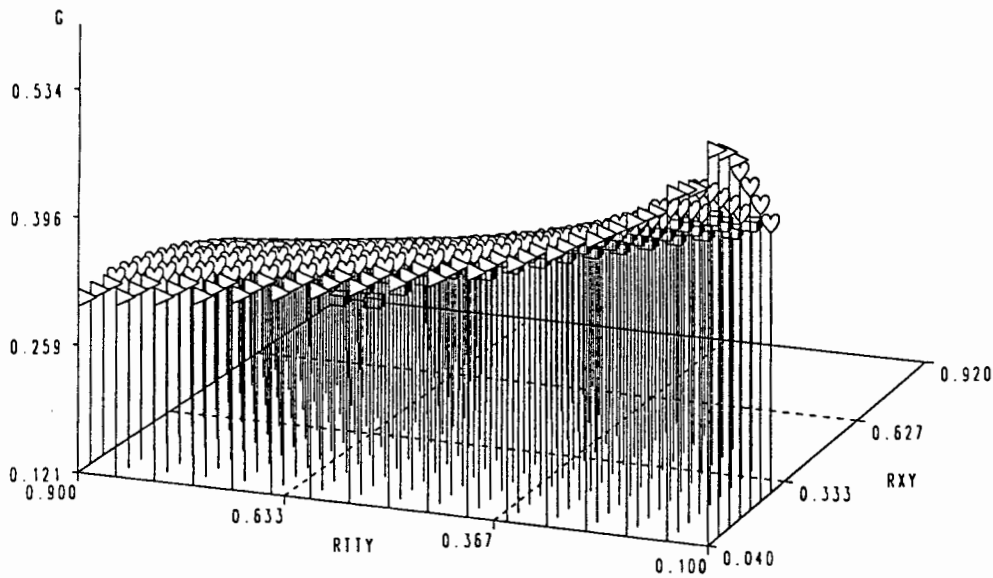


Figure 5.9:  $G = (\hat{\rho}_c / \sigma[\hat{\rho}_c]) / (\hat{\rho} / \sigma[\hat{\rho}]) = Z_c / Z$  as a function of  $\rho_{ttY}$ , the pre-correction correlation coefficient and sample size  $n$  when the reliability coefficient  $\rho_{ttY}$  is obtained from an independent data set;  $n$  is fixed at 120

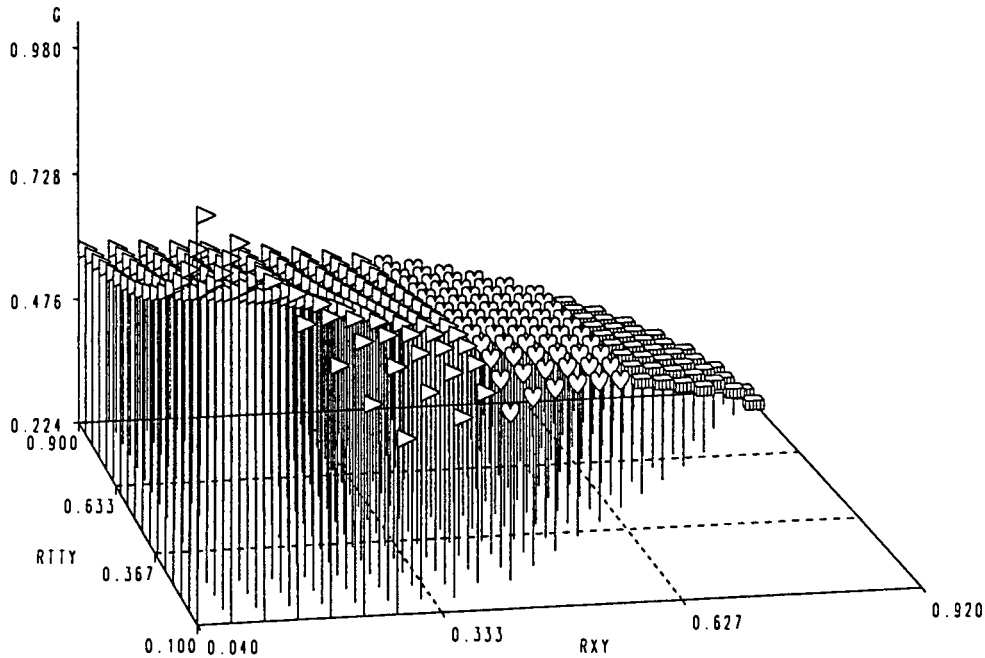


Figure 5.10:  $G = (\hat{\rho}_c/\sigma[\hat{\rho}_c])/(\hat{\rho}/\sigma[\hat{\rho}]) = Z_c/Z$  as a function of  $\rho_{ttY}$ , the pre-correction correlation coefficient and sample size  $n$  when the reliability coefficient  $\rho_{ttY}$  is obtained from an independent data set [rotated];  $n$  is fixed at 10

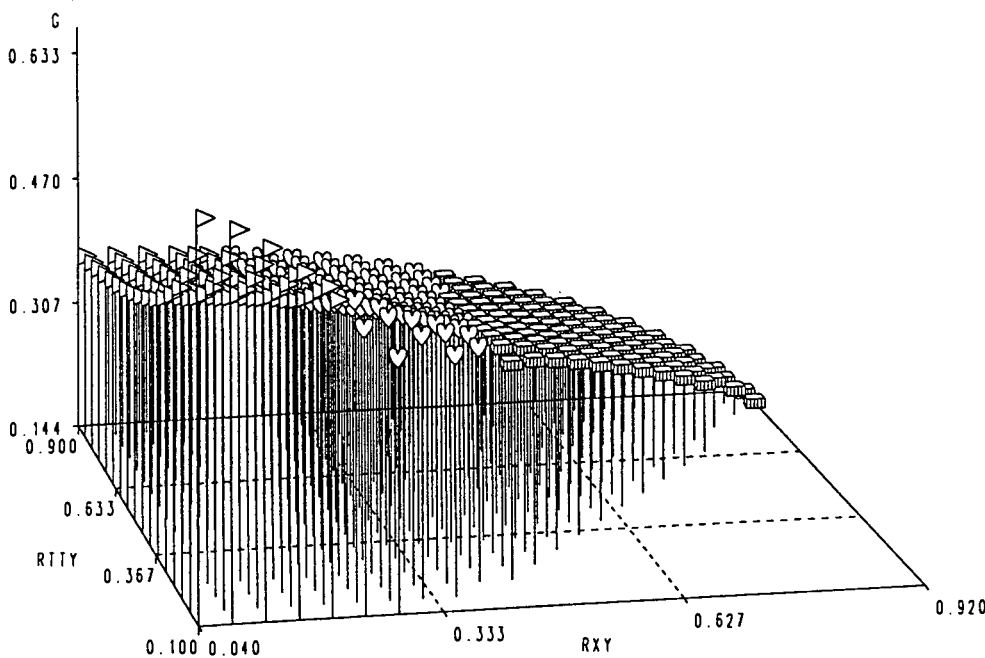


Figure 5.11:  $G = (\hat{\rho}_c/\sigma[\hat{\rho}_c])/(\hat{\rho}/\sigma[\hat{\rho}]) = Z_c/Z$  as a function of  $\rho_{ttY}$ , the pre-correction correlation coefficient and sample size  $n$  when the reliability coefficient  $\rho_{ttY}$  is obtained from an independent data set [rotated];  $n$  is fixed at 60

$\rho^*[X,Y]$ . The negative effect of the correction for criterion unreliability also seems to increase as  $\rho_{ttY}$  and  $\rho^*[X,Y]$  increases.

Bobko and Rieck [1980, p. 393] analytically reach a similar, although less detailed, conclusion:

If however  $r_{yy}$  is estimated independently of  $r_{yx}$ , B is inherently positive. In this case  $r_{xTy}$  is still larger than  $r_{xy}$  by a factor  $r_{yy}^{-1/2}$ . However,  $S.E.(r_{xTy})$  has increased by a greater factor, since  $S.E.(r_{xTy}) = (A+B)^{1/2} > A^{1/2} = r_{yy}^{-1/2}S.E.(r_{yx})$ . Thus,  $z_1 > z_2$  and it is recommended that the uncorrected coefficient be used for hypothesis testing, since it provides a more powerful test.

The latter recommendation should be challenged, however, since it implies that the one coefficient can serve as a logical substitute for the other. As was argued earlier a choice between  $\rho^*[X,Y]$  and  $\rho_c[X,Y]$  makes no sense since both constitute meaningful but qualitatively different, although related, quantities.

### 5.3.1.3 The Reliability Coefficient Obtained From The Same Data Set As $\rho^*[X,Y]$

The reaction of  $G = (\rho_c/\sigma[\rho_c])/(\rho^*/\sigma[\rho^*]) = Z_c/Z$  as a function of  $\rho_{ttY}$ , the pre-correction correlation coefficient and sample size  $n$  when the reliability coefficient  $\rho_{ttY}$  is obtained from the same sample as  $\rho^*[X,Y]$ , is graphically portrayed in Figures 5.12 - 5.15. An alternative perspective is again provided in Figure 5.16 and Figure 5.17 by rotating the plot towards the  $\rho^*[X,Y]$  axis. The behaviour of  $G$  when the standard error of the partially disattenuated correlation is calculated via the Bobko and Rieck [1980] Case I expression, is very similar to its behaviour under Case II conditions. All the conclusions that were deduced under paragraph 5.3.1.2 thus also apply to Case I conditions. These findings seems to be in line with the thinking of Bobko and Rieck [1980, p. 393], while at the same time offering a significantly more detailed and precise description of the effect of correcting the correlation coefficient for attenuation:

Finally, if  $r_{yy}$  and  $r_{xy}$  are estimated from the same sample, then the ordinal relationship between  $S.E.(r_{xTy}) = (A+B-C)^{1/2}$  and  $A^{1/2} = r_{yy}^{-1/2}S.E.(r_{xy})$  is not clear because  $(B-C)$  may be either positive or negative. Given that no interpretable pattern regarding the sign of  $(B-C)$  has been found, it is recommended that the uncorrected coefficient be tested.

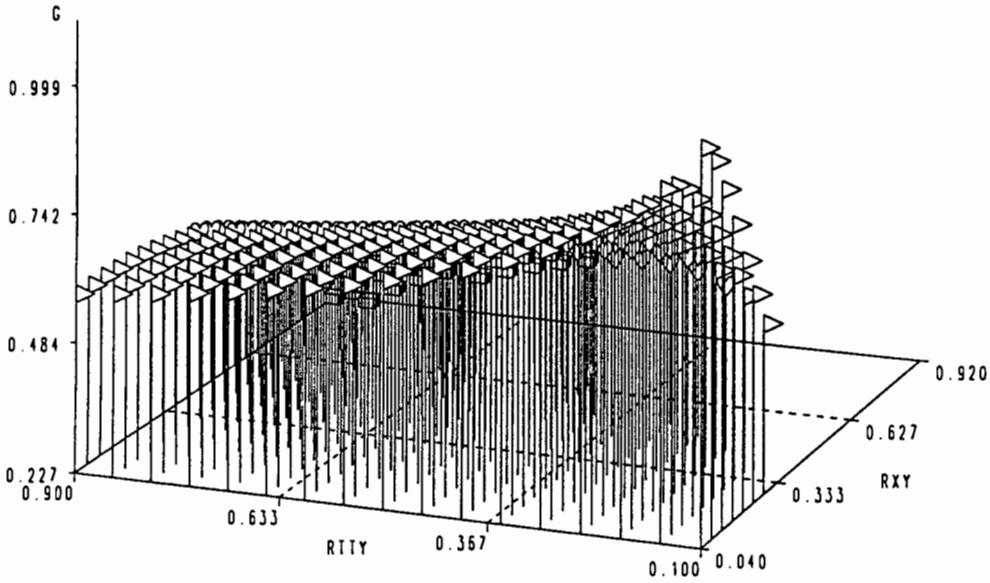


Figure 5.12:  $G = (\hat{\rho}_c/\sigma[\hat{\rho}_c]) / (\hat{\rho}/\sigma[\hat{\rho}]) = Z_c/Z$  as a function of  $\rho_{ttY}$ , the pre-correction correlation coefficient and sample size  $n$  when the reliability coefficient  $\rho_{ttY}$  is obtained from the same data set;  $n$  is fixed at 10

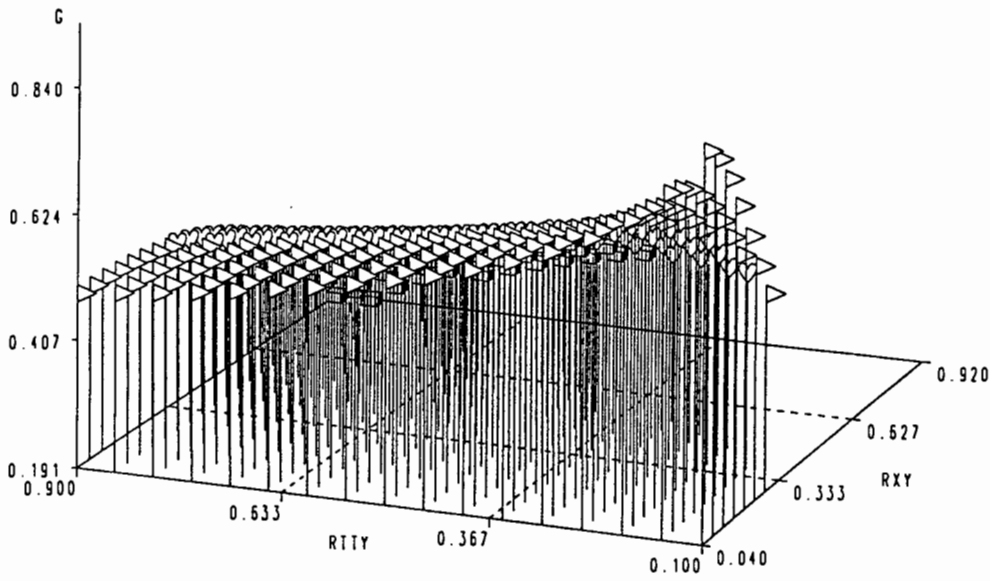


Figure 5.13:  $G = (\hat{\rho}_c/\sigma[\hat{\rho}_c]) / (\hat{\rho}/\sigma[\hat{\rho}]) = Z_c/Z$  as a function of  $\rho_{ttY}$ , the pre-correction correlation coefficient and sample size  $n$  when the reliability coefficient  $\rho_{ttY}$  is obtained from an independent data set;  $n$  is fixed at 20

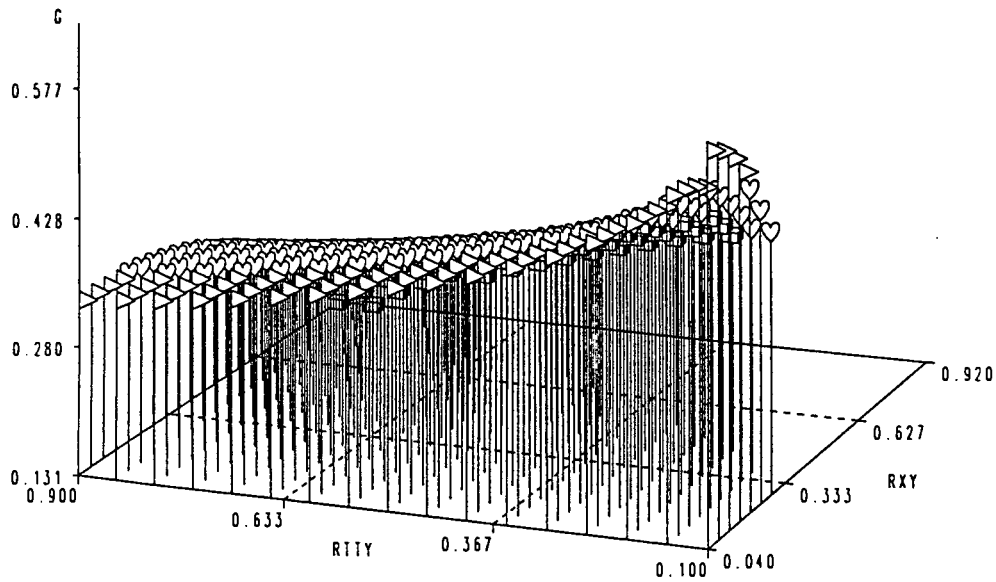


Figure 5.14:  $G = (\hat{\rho}_c / \sigma[\hat{\rho}_c]) / (\hat{\rho} / \sigma[\hat{\rho}]) = Z_c / Z$  as a function of  $\rho_{ttY}$ , the pre-correction correlation coefficient and sample size  $n$  when the reliability coefficient  $\rho_{ttY}$  is obtained from an independent data set;  $n$  is fixed at 90

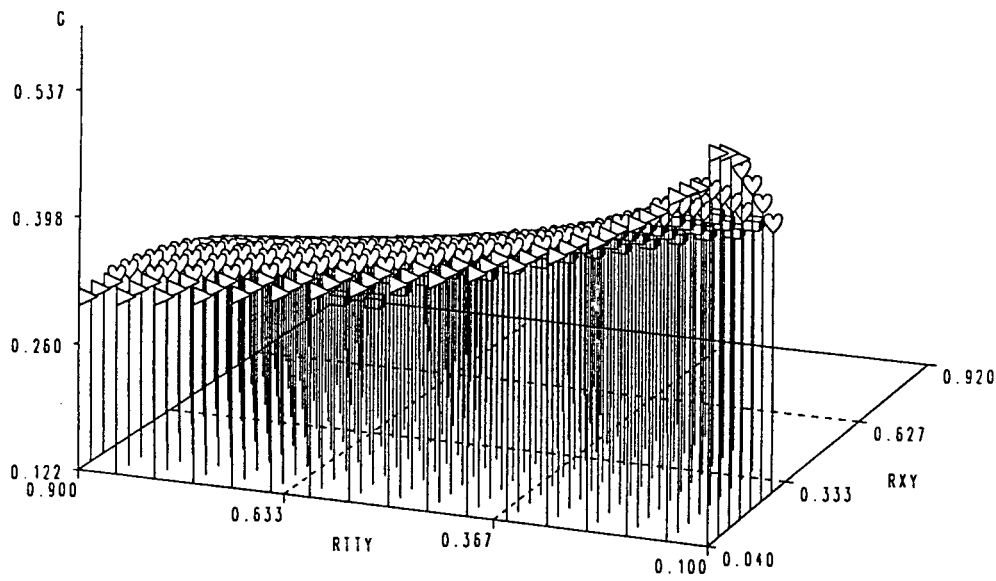


Figure 5.15:  $G = (\hat{\rho}_c / \sigma[\hat{\rho}_c]) / (\hat{\rho} / \sigma[\hat{\rho}]) = Z_c / Z$  as a function of  $\rho_{ttY}$ , the pre-correction correlation coefficient and sample size  $n$  when the reliability coefficient  $\rho_{ttY}$  is obtained from an independent data set;  $n$  is fixed at 120

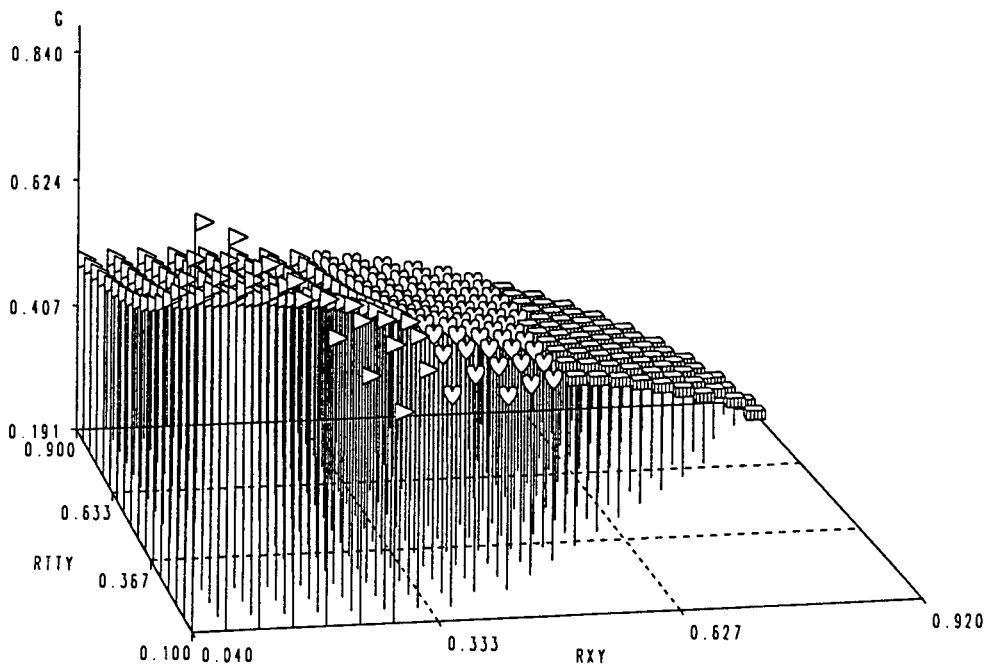


Figure 5.16:  $G = (\hat{\rho}_c/\sigma[\hat{\rho}_c]) / (\hat{\rho}/\sigma[\hat{\rho}]) = Z_c/Z$  as a function of  $\rho_{ttY}$ , the pre-correction correlation coefficient and sample size  $n$  when the reliability coefficient  $\rho_{ttY}$  is obtained from an independent data set [rotated];  $n$  is fixed at 20

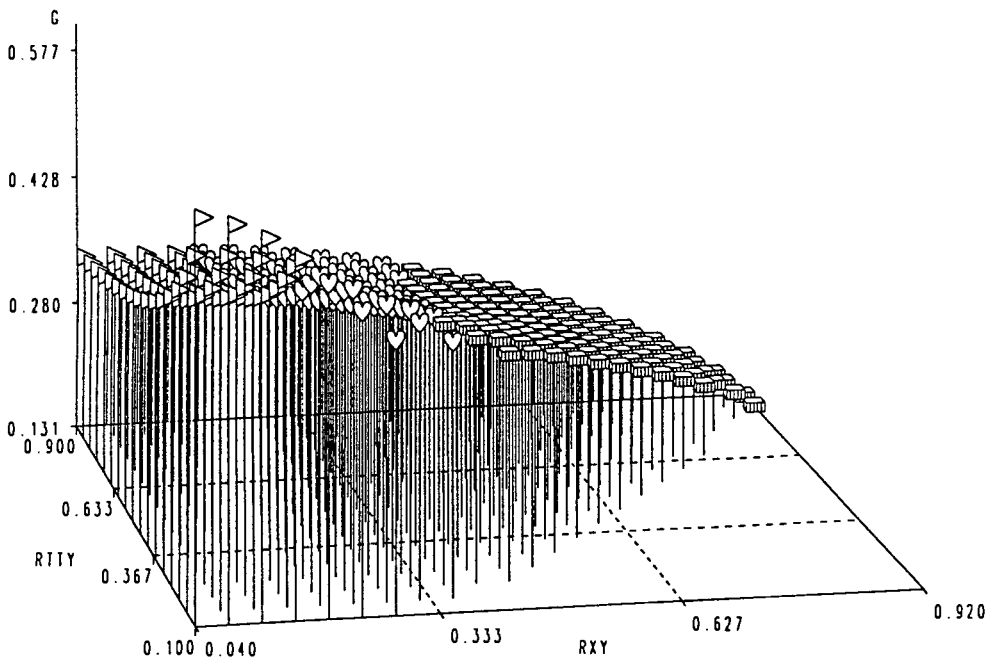


Figure 5.17:  $G = (\hat{\rho}_c/\sigma[\hat{\rho}_c]) / (\hat{\rho}/\sigma[\hat{\rho}]) = Z_c/Z$  as a function of  $\rho_{ttY}$ , the pre-correction correlation coefficient and sample size  $n$  when the reliability coefficient  $\rho_{ttY}$  is obtained from an independent data set [rotated];  $n$  is fixed at 90



### 5.3.2 Effect Of Corrections For Restriction Of Range To The Pearson Correlation Coefficient On The Empirically Derived Exceedence Probabilities $\alpha_b$

#### 5.3.2.1 Case 2 [Case A] Restriction Of Range

Figures 5.18 - 5.21 depict the reaction of  $G$  to changes in the initial, obtained correlation coefficient  $\rho^{\wedge}[x,y]$ , the ratio  $K = \sigma[X]/\sigma[x]$  [and by implication, therefore, the selection ratio  $[\phi]$  since  $K = [1+(z_c\lambda/\phi)-(\lambda/\phi)^2]^{-1}$ ] and sample size  $n$  when the obtained correlation had been attenuated through Case 2 [Case A] selection. Figures 5.18 - 5.21 reveal that  $G \geq 1$  for all combinations of  $\rho^{\wedge}[x,y]$ ,  $K$  and  $n$ . For  $K = 1$ ,  $G$  logically equals unity since  $\rho^{\wedge}[x,y] = \rho^{\wedge}[X,Y]$  and  $\sigma(\rho^{\wedge}[x,y]) = \sigma(\rho^{\wedge}[X,Y])$ . It thus follows that for all combinations of  $n$ ,  $\rho^{\wedge}[x,y]$  and  $K > 1$ , the probability of observing  $\rho^{\wedge}[x,y]$  conditional on  $H_0: \rho[x,y] = 0$  [i.e.  $\alpha_b$ ] is systematically bigger than the probability of observing the corresponding corrected correlation coefficient conditional on the hypothesis that  $\rho[X,Y] = 0$ . Figures 5.18 - 5.21 furthermore demonstrate that the effect of the Case 2 [Case A] correction for restriction of range on  $\alpha_b$  intensifies curvilinearly as  $\rho^{\wedge}[x,y]$  and  $K$  increases. In addition, the decrease in  $\alpha_b$  produced by the correction for restriction of range has the effect of changing some insignificant uncorrected correlations into significant corrected correlations [see Table 5.2]. This effect seems to be relatively sensitive to sample size. At  $n \geq 120$  Case 2 [Case A] restriction of range corrections to the correlation coefficient no longer have the effect of changing the significance test outcome [given  $\alpha_c = 0.05$ ]. Figures 5.18 - 5.21 suggest the change in decision outcome to occur in a specific region on the  $\rho^{\wedge}[X,Y] - K$  space as a function of sample size. As sample size increases, the region shifts to lower values of  $\rho^{\wedge}[X,Y]$  but simultaneously also higher values of  $K$  [i.e. smaller/stricter selection ratios]. These findings are in agreement with the analytical conclusions reached by Bobko and Rieck [1980, p. 393] but simultaneously offering a significantly more detailed and precise description of the effect of correcting the correlation coefficient for Case 2 [Case A] restriction of range:

Regarding the correlation corrected for restriction of range, Moran's test compares  $z_3 = r_{\text{range}}/\text{S.E.}(r_{\text{range}})$  to a standard normal distribution. Again letting  $z_1 = r/\text{S.E.}(r)$ , it is easily shown [See Equation 5.2] that  $z_3/z_1 = [1+r^2(K-1)] > 1$  if  $K > 1$ . Thus the Moran test of a validity coefficient corrected for restriction in range is asymptotically more powerful than the test of the uncorrected coefficient.

$$G = (\rho^{\wedge}_c/\sigma[\rho^{\wedge}_c]) / (\rho^{\wedge}/\sigma[\rho^{\wedge}]) \\ = (\rho^{\wedge}_c/\sigma[\rho^{\wedge}_c])(\sigma[\rho^{\wedge}]/\rho^{\wedge}) \text{-----} 5.2.1$$

Substituting the expression for the correlation corrected for Case 2 [Case A] restriction of range [Equation 3.4] in Equation 5.2.1:

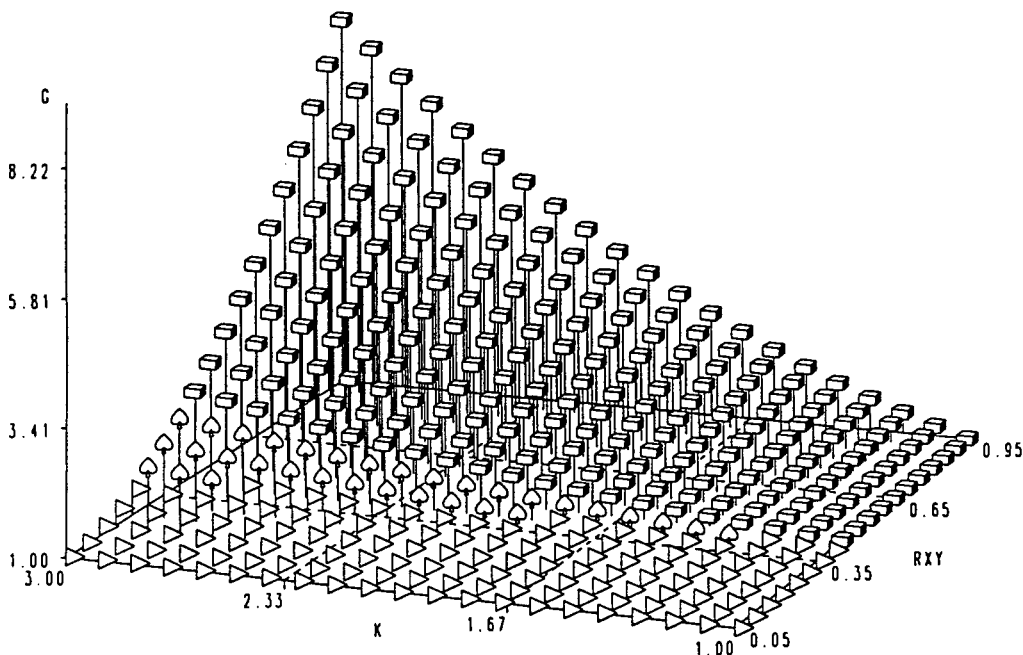


Figure 5.18:  $G = (\hat{\rho}_c / \sigma[\hat{\rho}_c]) / (\hat{\rho} / \sigma[\hat{\rho}]) = Z_c / Z$  under Case 2 [Case A] restriction of range as a function of  $\rho_{\tau Y}$  and the pre-correction correlation coefficient for  $n$  fixed at 10

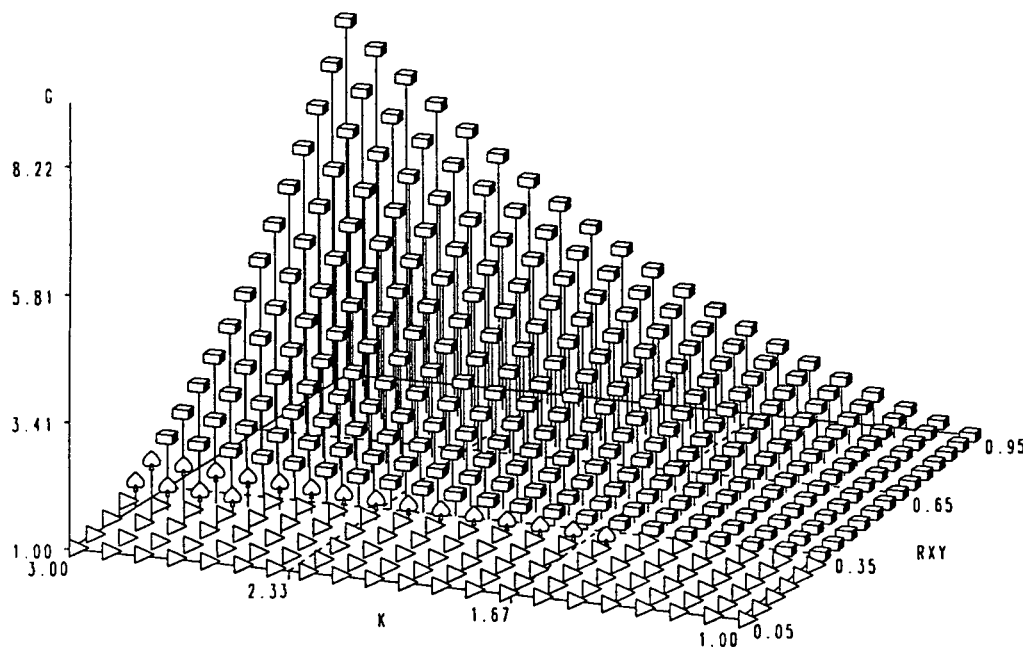


Figure 5.19:  $G = (\hat{\rho}_c / \sigma[\hat{\rho}_c]) / (\hat{\rho} / \sigma[\hat{\rho}]) = Z_c / Z$  under Case 2 [Case A] restriction of range as a function of  $\rho_{\tau Y}$  and the pre-correction correlation coefficient for  $n$  fixed at 20

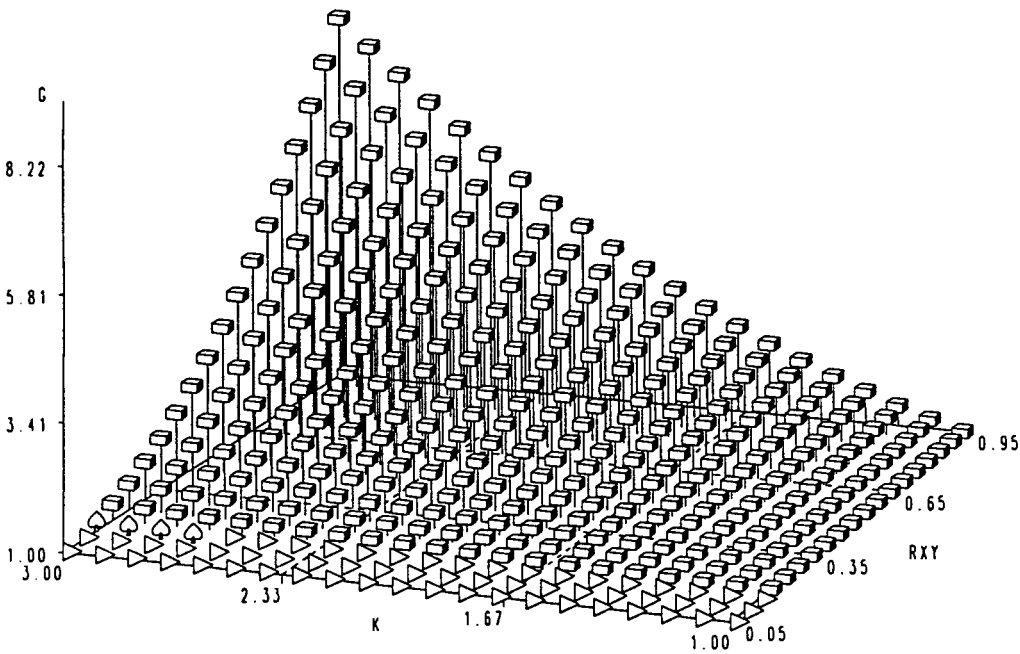


Figure 5.20:  $G = (\hat{\rho}_c / \sigma[\hat{\rho}_c]) / (\hat{\rho} / \sigma[\hat{\rho}]) = Z_c / Z$  under Case 2 [Case A] restriction of range as a function of  $\rho_{\tau\tau Y}$  and the pre-correction correlation coefficient for  $n$  fixed at 90

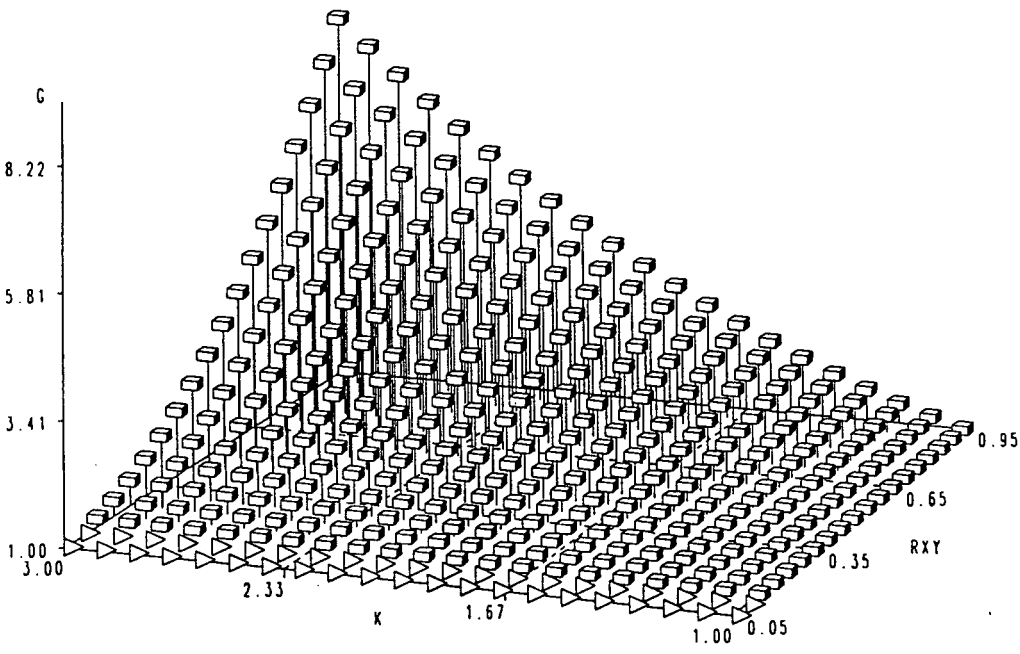


Figure 5.21:  $G = (\hat{\rho}_c / \sigma[\hat{\rho}_c]) / (\hat{\rho} / \sigma[\hat{\rho}]) = Z_c / Z$  under Case 2 [Case A] restriction of range, as a function of  $\rho_{\tau\tau Y}$  and the pre-correction correlation coefficient for  $n$  fixed at 120

$$G = (\rho^{\wedge}K/(1-\rho^{\wedge 2}+\rho^{\wedge 2}K^2)^{1/2})(1/\sigma[\rho^{\wedge}c])(\sigma[\rho^{\wedge}]/\rho^{\wedge}) \text{-----} 5.2.2$$

Canceling out the term  $\rho^{\wedge}$  and substituting Equation 4.35 and Equation 3.4 in Equation 5.2.2:

$$\begin{aligned} G &= (K/(1-\rho^{\wedge 2}+\rho^{\wedge 2}K^2)^{1/2})((1-\rho^{\wedge 2}+\rho^{\wedge 2}K^2)^{3/2}/K)(1/\sigma[\rho^{\wedge}]) (\sigma[\rho^{\wedge}]) \\ &= (1-\rho^{\wedge 2}+\rho^{\wedge 2}K^2) \text{-----} 5.2 \end{aligned}$$

Therefore it follows from Equation 5.2 that:

For  $K > 1$ ,  $G > 1$

### 5.3.2.2 Case 3 [Case C] Restriction Of Range

The behaviour of  $G$  under changes of  $K$ ,  $n$ ,  $\rho^{\wedge}[x,y]$ ,  $\rho^{\wedge}[x,z]$  and  $\rho^{\wedge}[y,z]$ , when the obtained correlation had been attenuated by selection on a third variable  $Z$ , is depicted in Figures 5.22 - 5.27. An examination of Figures 5.22 - 5.27 firstly confirms the logically necessary conclusion that for  $K = 1$ , for all permissible combinations of values of  $\rho^{\wedge}[x,y]$ ,  $\rho^{\wedge}[x,z]$ ,  $\rho^{\wedge}[y,z]$  and  $n$ ,  $G = 1$ . Figures 5.22 - 5.27 furthermore indicate that for  $K > 1$ ,  $G$  reacts to changes in  $K$ ,  $n$ ,  $\rho^{\wedge}[x,y]$ ,  $\rho^{\wedge}[x,z]$  and  $\rho^{\wedge}[y,z]$  with values greater than and less than 1. Correcting the correlation coefficient for Case C restriction of range thus increases the conditional probability  $\alpha_b = P[\rho^{\wedge}_c \geq \rho^{\wedge}_{cb} | \rho^{\wedge}_c = 0]$  relative to  $\alpha_b = P[\rho^{\wedge} \geq \rho^{\wedge}_b | \rho^{\wedge} = 0]$  at some combinations of values of  $n$ ,  $K$  [ $K > 1$ ],  $\rho^{\wedge}[x,y]$ ,  $\rho^{\wedge}[x,z]$  and  $\rho^{\wedge}[y,z]$  [thereby reducing the chances of rejecting  $H_0$ ] while it decreases  $\alpha_b$  associated with the corrected coefficient relative to  $\alpha_b$  associated with  $\rho^{\wedge}$  in other regions of the six dimensional space [thereby increasing the chances of rejecting  $H_0$ ]. These changes in  $\alpha_b$  produced by the correction for restriction of range has the effect of changing some insignificant uncorrected correlations into significant corrected correlations and vice versa [see Table 5.2].

The region of change from insignificance to significance seems:

- to occur at the lower end of the  $\rho^{\wedge}[x,y]$  axis for moderate to high  $\rho^{\wedge}[x,z]$  values;
- to shift towards the upper end of the  $\rho^{\wedge}[x,z]$  axis as  $\rho^{\wedge}[y,z]$  increases;
- to spread towards zero on the  $\rho^{\wedge}[x,y]$  axis as  $K$  increases; and
- to diminish as  $n$  increases.

In contrast, the region of change from significance to insignificance seems:

- to occur at the lower end of the  $\rho^{\wedge}[x,z]$  axis for moderate to high  $\rho^{\wedge}[x,y]$  values;
- to shift towards the upper end of the  $\rho^{\wedge}[x,y]$  axis as  $\rho^{\wedge}[y,z]$  increases;
- to spread towards zero on the  $\rho^{\wedge}[x,z]$  axis as  $K$  decreases; and

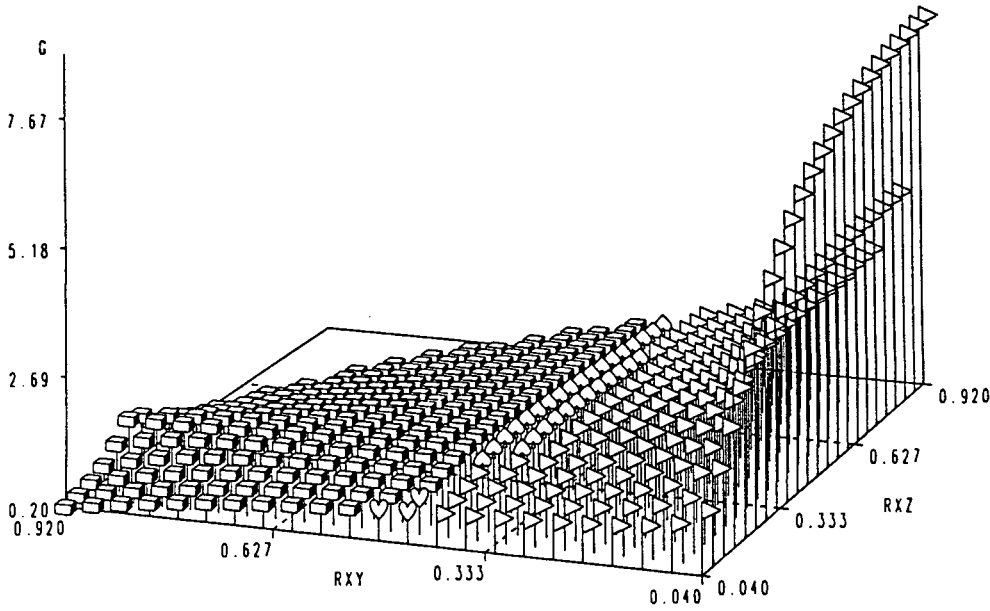


Figure 5.22:  $G = (\hat{\rho}_c / \sigma[\hat{\rho}_c]) / (\hat{\rho} / \sigma[\hat{\rho}]) = Z_c / Z$  under Case 3[i] [Case C[i]] restriction of range as a function of  $\hat{\rho}[x,y]$  and  $\hat{\rho}[x,z]$  for  $n$  fixed at 10,  $K$  fixed at 2 and  $\hat{\rho}[y,z]$  fixed at 0.25

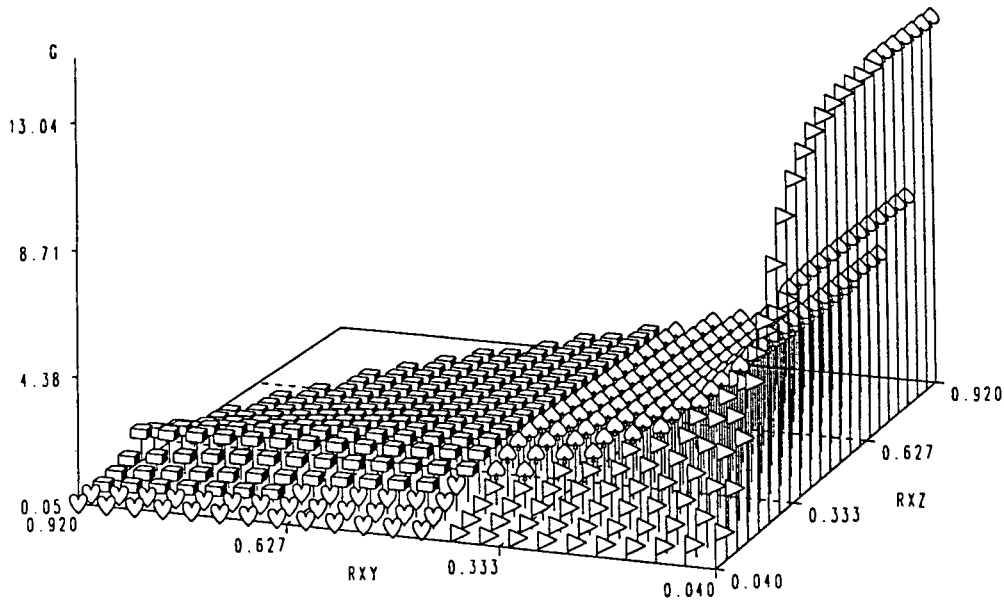


Figure 5.23:  $G = (\hat{\rho}_c / \sigma[\hat{\rho}_c]) / (\hat{\rho} / \sigma[\hat{\rho}]) = Z_c / Z$  under Case 3[i] [Case C[i]] restriction of range as a function of  $\hat{\rho}[x,y]$  and  $\hat{\rho}[x,z]$  for  $n$  fixed at 10,  $K$  fixed at 4 and  $\hat{\rho}[y,z]$  fixed at 0.25



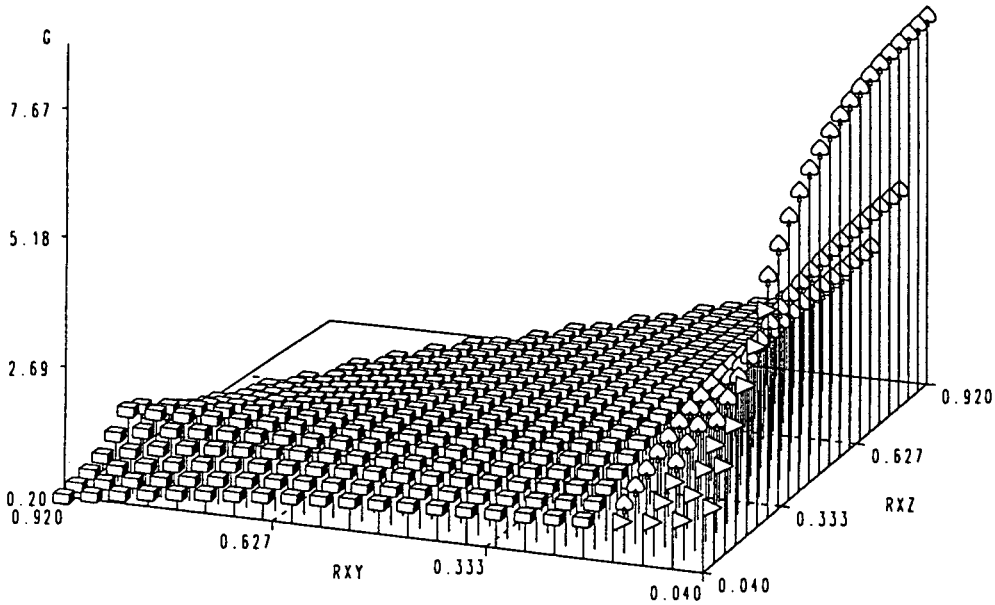


Figure 5.24:  $G = (\hat{\rho}_c/\sigma[\hat{\rho}_c])/(\hat{\rho}/\sigma[\hat{\rho}]) = Z_c/Z$  under Case 3[i] [Case C[i]] restriction of range as a function of  $\hat{\rho}[x,y]$  and  $\hat{\rho}[x,z]$  for  $n$  fixed at 90,  $K$  fixed at 2 and  $\hat{\rho}[y,z]$  fixed at 0.25

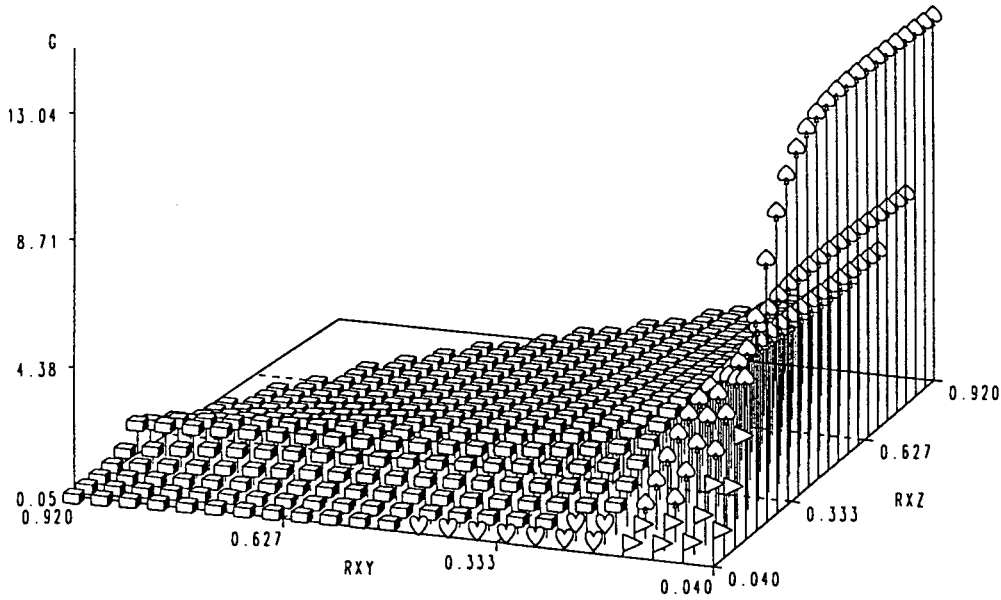


Figure 5.25:  $G = (\hat{\rho}_c/\sigma[\hat{\rho}_c])/(\hat{\rho}/\sigma[\hat{\rho}]) = Z_c/Z$  under Case 3[i] [Case C[i]] restriction of range as a function of  $\hat{\rho}[x,y]$  and  $\hat{\rho}[x,z]$  for  $n$  fixed at 90,  $K$  fixed at 2 and  $\hat{\rho}[y,z]$  fixed at 0.25



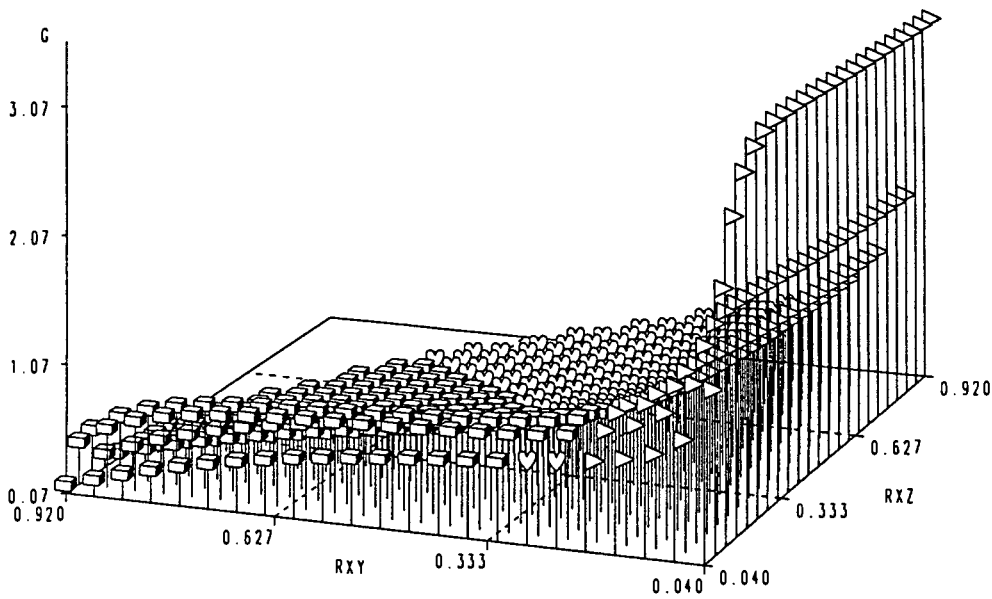


Figure 5.26:  $G = (\hat{\rho}_c/\sigma[\hat{\rho}_c])/(\hat{\rho}/\sigma[\hat{\rho}]) = Z_c/Z$  under Case 3[i] [Case C[i]] restriction of range as a function of  $\hat{\rho}[x,y]$  and  $\hat{\rho}[x,z]$  for  $n$  fixed at 60,  $K$  fixed at 4 and  $\hat{\rho}[y,z]$  fixed at 0.10

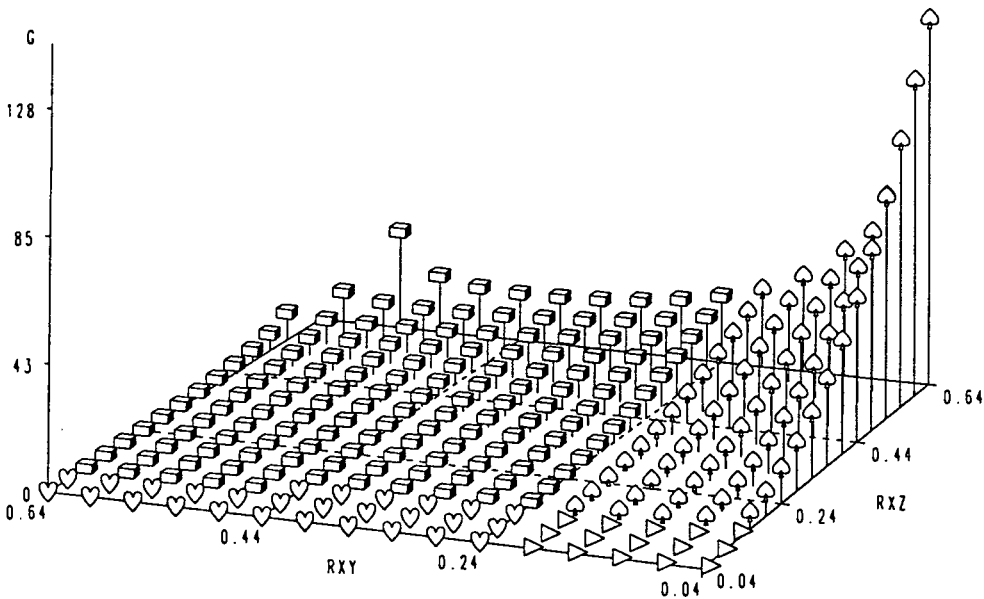


Figure 5.27:  $G = (\hat{\rho}_c/\sigma[\hat{\rho}_c])/(\hat{\rho}/\sigma[\hat{\rho}]) = Z_c/Z$  under Case 3[i] [Case C[i]] restriction of range as a function of  $\hat{\rho}[x,y]$  and  $\hat{\rho}[x,z]$  for  $n$  fixed at 60,  $K$  fixed at 4 and  $\hat{\rho}[y,z]$  fixed at 0.75

- to diminish as  $n$  increases.

The effect of corrections for Case 3[i] [Case C[i]] restriction of range thus seems somewhat more complex than the effect of corrections for Case 2 [Case A] restriction of range. Inspection of Figures 5.22 - 5.27 suggests the following additional specific conclusions on the effect of corrections for Case C restriction of range to be appropriate:

- $G$  increases as  $K$  increases [i.e. as the selection ratio decreases or selection becomes more severe];
- for any  $K > 1$ ,  $G$  increases as the correlation of the selection variable  $Z$  with the predictor  $X$  increases;
- for any  $K > 1$ ,  $G$  increases as the correlation of the selection variable  $Z$  with the criterion  $Y$  increases; and
- for any  $K > 1$ , higher  $G$  values are found at extreme low values of  $\rho^{[x,y]}$ ;

### 5.3.3 Effect Of Corrections For Criterion Unreliability And Restriction Of Range To The Pearson Correlation Coefficient On The Empirically Derived Exceedence Probabilities $\alpha_b$

The reaction of  $G = (\rho^c_c / \sigma[\rho^c_c]) / (\rho^{\wedge} / \sigma[\rho^{\wedge}])$  to changes in  $\rho^{[x,y]}$ ,  $K$ ,  $n$  and  $\rho_{ttY}$  when correcting the obtained correlation coefficient for criterion unreliability and [Case 2 [Case A]] restriction of range simultaneously is depicted in Figures 5.28 - 5.35. Figures 5.28 - 5.35 indicate  $G > 1$  for  $K > 1$  for all values of  $\rho^{[x,y]}$  and  $\rho_{ttY}$ . Figure 5.28 and Figure 5.32 indicate  $G < 1$  for  $K = 1$  in a small region defined by low values on both the  $\rho_{ttY}$  and  $\rho^{[x,y]}$  axes.

It thus follows that for all combinations of  $n$ ,  $\rho^{[x,y]}$ ,  $\rho_{ttY}$  and  $K > 1$  the probability of observing  $\rho^{[x,y]}$  conditional on  $H_0: \rho[x,y] = 0$  [i.e.  $\alpha_b$ ] is systematically bigger than the probability of observing the corresponding double corrected correlation coefficient conditional on the hypothesis that  $\rho[X,TY] = 0$ . Figures 5.28 - 5.35 furthermore demonstrate that the effect of the double correction for [Case 2 [Case A]] restriction of range and criterion unreliability on  $\alpha_b$  intensifies [i.e.  $G$  increases] as  $\rho^{[x,y]}$  and  $K$  increases and  $\rho_{ttY}$  decreases. In addition, the decrease in  $\alpha_b$  produced by the double correction has the effect of changing some insignificant uncorrected correlations into significant corrected correlations [see Table 5.2] when  $K > 1$ . For  $K > 1$  the opposite effect does not seem to occur. The aforementioned effect seems to be relatively sensitive to sample size. At  $n \geq 120$ , double corrections to the correlation coefficient no longer have the effect of changing the significance test outcome [given  $\alpha_c = 0.05$ ]. Figures 5.28 - 5.35 suggest the change in decision outcome to occur in the lower region of the  $\rho^{[x,y]}$  axis and the full range of  $\rho_{ttY}$  values as a function of sample size and  $K$ . As sample size

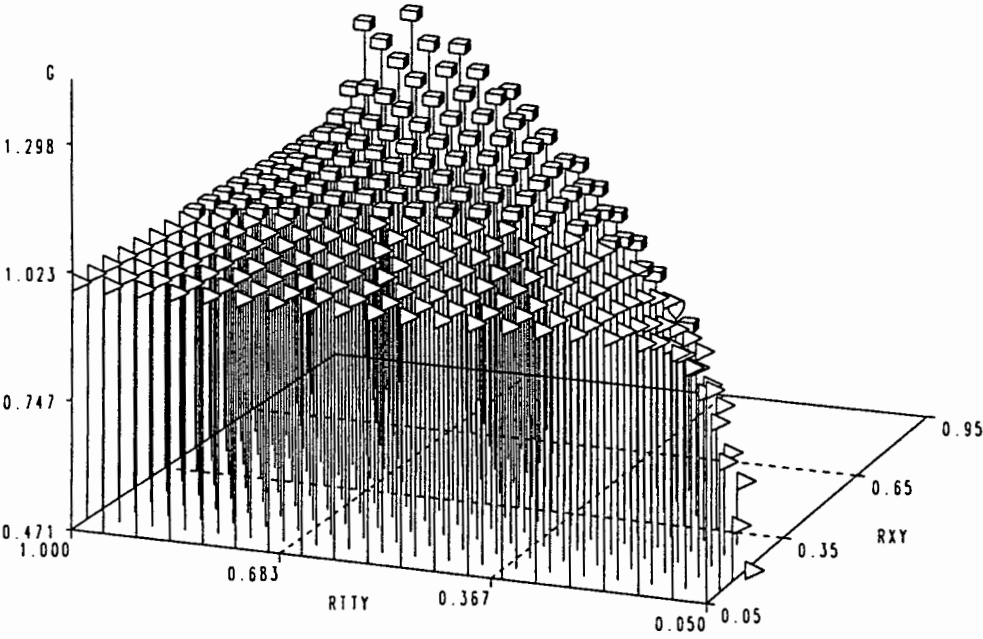


Figure 5.28:  $G = (\hat{\rho}_c / \sigma[\hat{\rho}_c]) / (\hat{\rho} / \sigma[\hat{\rho}])$  as a function of  $\hat{\rho}[x,y]$  and  $\rho_{ttY}$  when correcting the obtained correlation coefficient for criterion unreliability and [Case 2 [Case A]] restriction of range simultaneously;  $n$  is fixed at 10 and  $K$  is fixed at 1

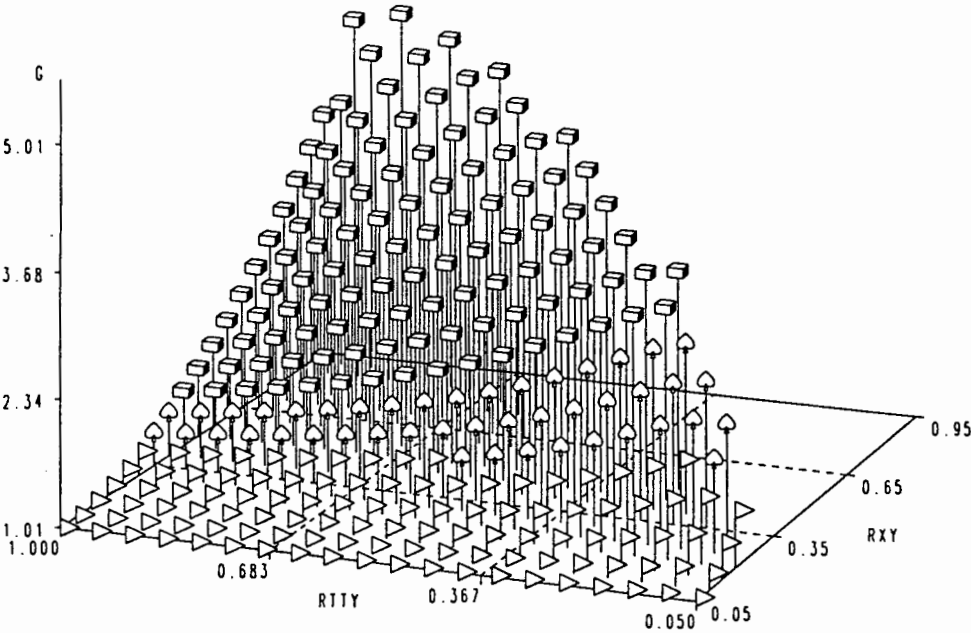


Figure 5.29:  $G = (\hat{\rho}_c / \sigma[\hat{\rho}_c]) / (\hat{\rho} / \sigma[\hat{\rho}])$  as a function of  $\hat{\rho}[x,y]$  and  $\rho_{ttY}$  when correcting the obtained correlation coefficient for criterion unreliability and [Case 2 [Case A]] restriction of range simultaneously;  $n$  is fixed at 10 and  $K$  is fixed at 2

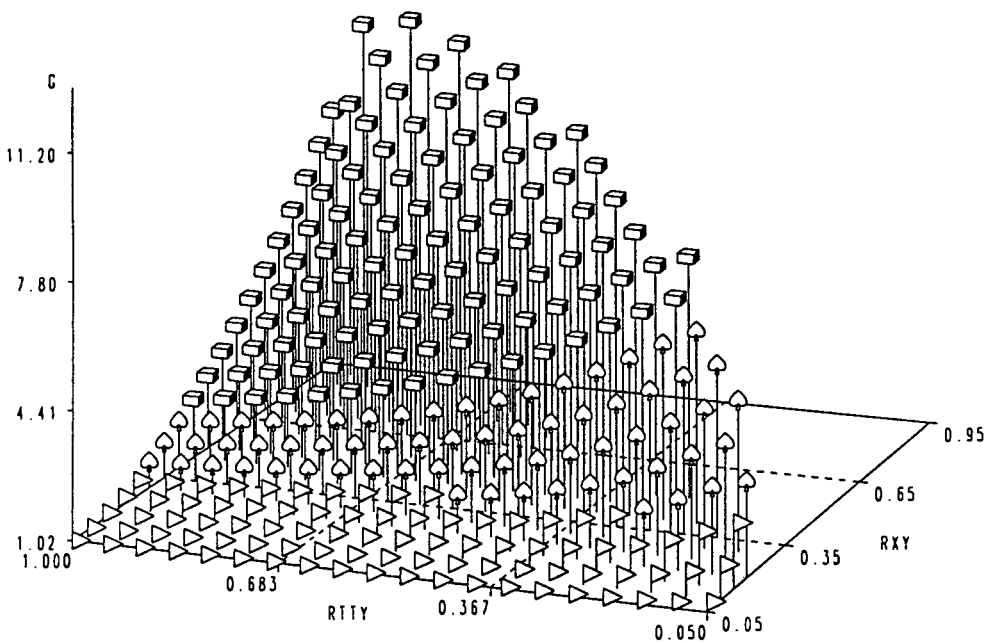


Figure 5.30:  $G = (\hat{\rho}_c / \sigma[\hat{\rho}_c]) / (\hat{\rho} / \sigma[\hat{\rho}])$  as a function of  $\hat{\rho}[x,y]$  and  $\rho_{ttY}$  when correcting the obtained correlation coefficient for criterion unreliability and [Case 2 [Case A]] restriction of range simultaneously;  $n$  is fixed at 10 and  $K$  is fixed at 3

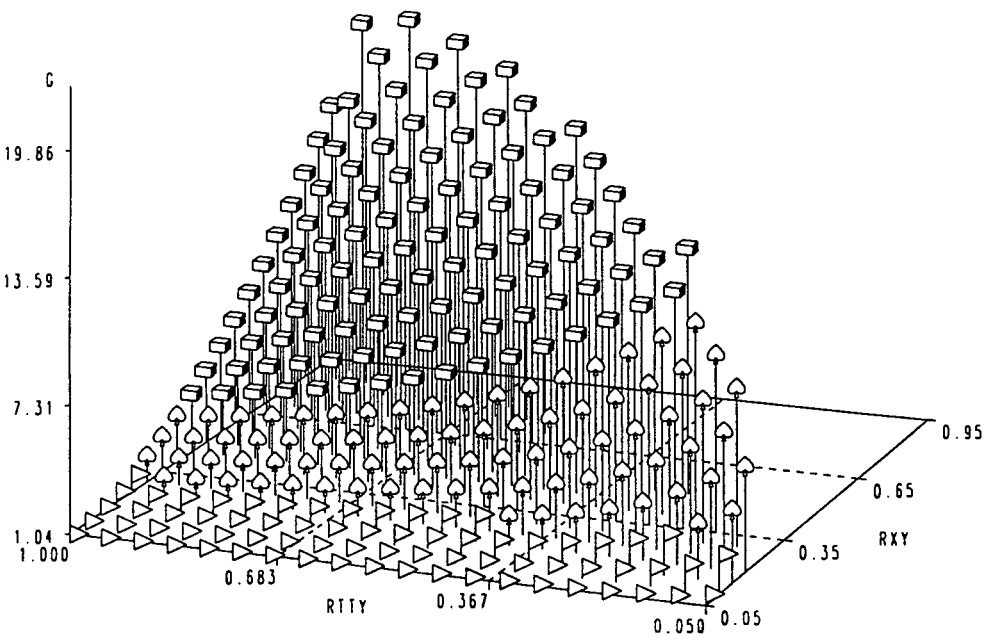


Figure 5.31:  $G = (\hat{\rho}_c / \sigma[\hat{\rho}_c]) / (\hat{\rho} / \sigma[\hat{\rho}])$  as a function of  $\hat{\rho}[x,y]$  and  $\rho_{ttY}$  when correcting the obtained correlation coefficient for criterion unreliability and [Case 2 [Case A]] restriction of range simultaneously;  $n$  is fixed at 10 and  $K$  is fixed at 4

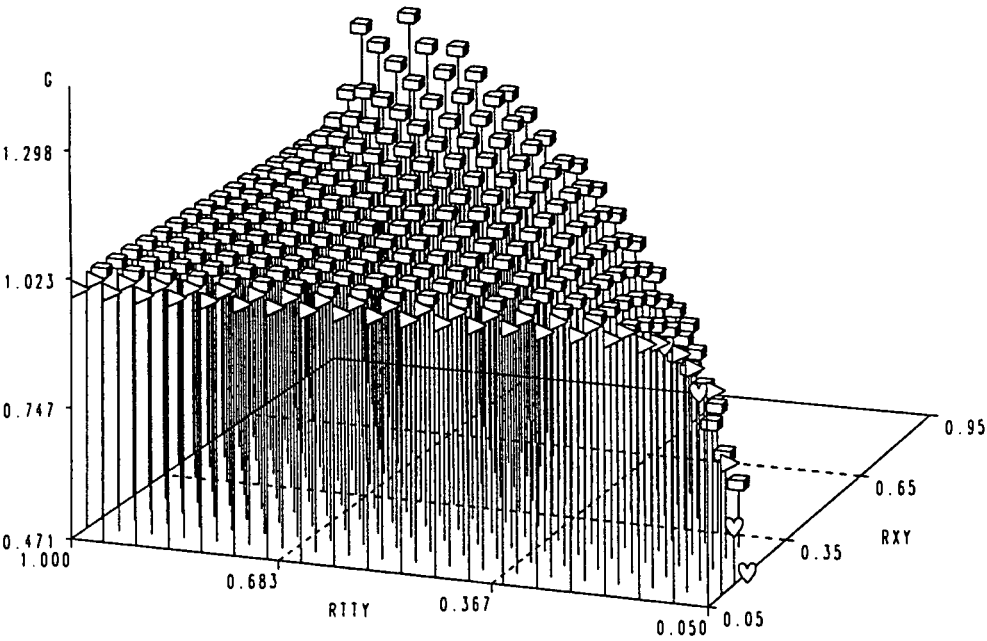


Figure 5.32:  $G = (\hat{\rho}_c / \sigma[\hat{\rho}_c]) / (\hat{\rho} / \sigma[\hat{\rho}])$  as a function of  $\hat{\rho}[x,y]$  and  $\rho_{ttY}$  when correcting the obtained correlation coefficient for criterion unreliability and [Case 2 [Case A]] restriction of range simultaneously;  $n$  is fixed at 120 and  $K$  is fixed at 1

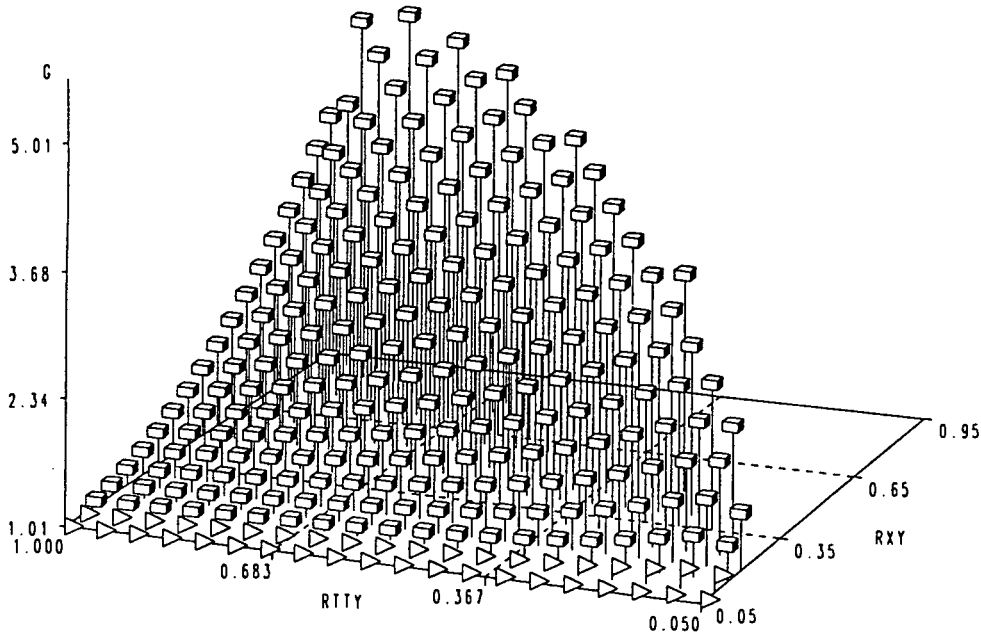


Figure 5.33:  $G = (\hat{\rho}_c / \sigma[\hat{\rho}_c]) / (\hat{\rho} / \sigma[\hat{\rho}])$  as a function of  $\hat{\rho}[x,y]$  and  $\rho_{ttY}$  when correcting the obtained correlation coefficient for criterion unreliability and [Case 2 [Case A]] restriction of range simultaneously;  $n$  is fixed at 120 and  $K$  is fixed at 2



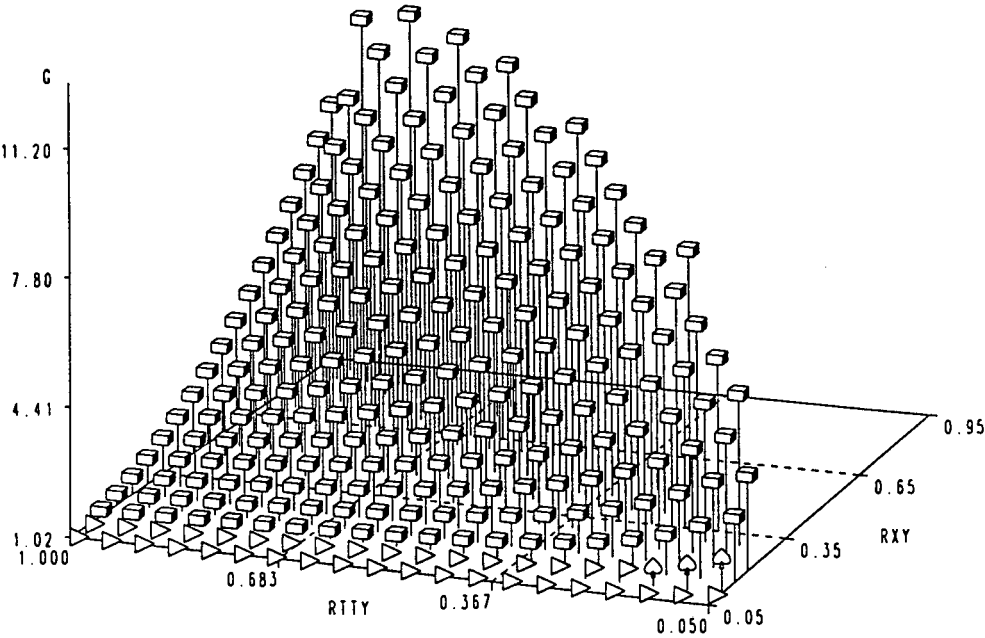


Figure 5.34:  $G = (\hat{\rho}_c / \sigma[\hat{\rho}_c]) / (\hat{\rho} / \sigma[\hat{\rho}])$  as a function of  $\hat{\rho}[x,y]$  and  $\rho_{ttY}$  when correcting the obtained correlation coefficient for criterion unreliability and [Case 2 [Case A]] restriction of range simultaneously;  $n$  is fixed at 120 and  $K$  is fixed at 3

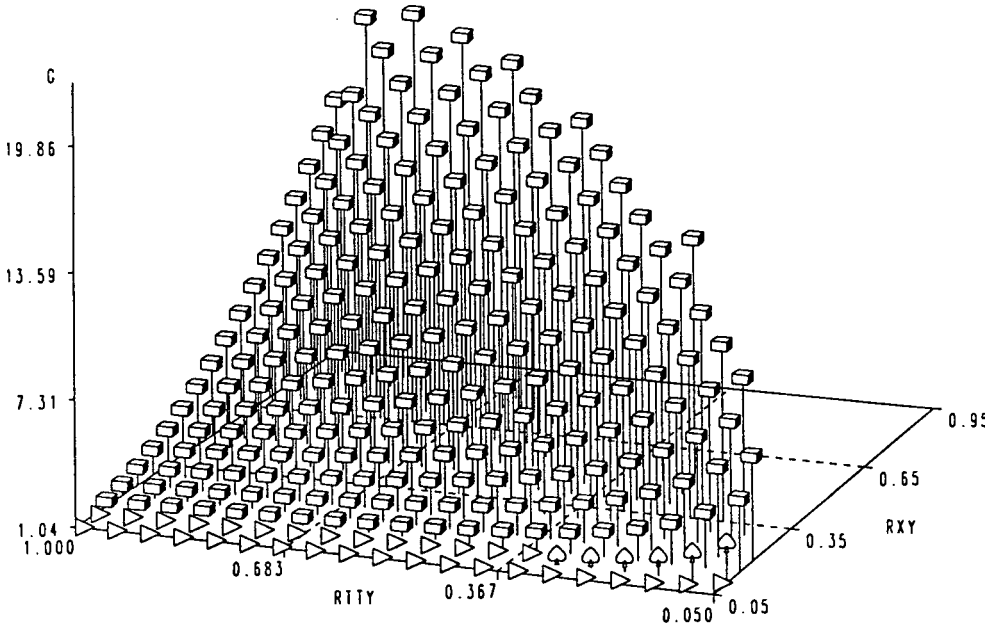


Figure 5.35:  $G = (\hat{\rho}_c / \sigma[\hat{\rho}_c]) / (\hat{\rho} / \sigma[\hat{\rho}])$  as a function of  $\hat{\rho}[x,y]$  and  $\rho_{ttY}$  when correcting the obtained correlation coefficient for criterion unreliability and [Case 2 [Case A]] restriction of range simultaneously,  $n$  is fixed at 120,  $K$  is fixed at 4



increases the region shifts to lower values of  $\rho^*[x,y]$ . As sample size increases the spread of the region also seems to be restricted more and more to the lower regions of the  $r_{ttY}$  axis. The region tends to increase at higher values of  $K$  [i.e. smaller/stricter selection ratios].

#### 5.4 EFFECT OF CORRECTIONS FOR ATTENUATION AND/OR RESTRICTION OF RANGE TO THE PEARSON CORRELATION COEFFICIENT ON THE A PRIORI PROBABILITIES $[1-\beta]$ OR POWER OF THE TESTS OF $H_0$ .

Previously it was argued that the effect of corrections to the correlation coefficient for the attenuating effect of restriction of range and/or criterion unreliability on statistical power would depend on how the ratio  $\rho^*/\sigma[\rho^*]$  compares to the ratio  $\rho_c^*/\sigma[\rho_c^*]$ . The critical ratio is thus  $(\rho^*/\sigma[\rho^*]) / (\rho_c^*/\sigma[\rho_c^*])$ . Specifically it was argued that if  $(\rho^*/\sigma[\rho^*]) / (\rho_c^*/\sigma[\rho_c^*])$  should equal unity, the power of the test of the hypothesis  $H_0: \rho = 0$  would equal the power of the test of the hypothesis  $H_0: \rho_c = 0$ . If, however,  $(\rho^*/\sigma[\rho^*]) / (\rho_c^*/\sigma[\rho_c^*])$  should drop below 1, power would be positively affected since it would imply that the critical cutoff under  $H_0$  would translate to a more extreme standard score under  $H_a$  for the corrected correlation coefficient than for the uncorrected coefficient. It thus follows that if  $\rho_c = \rho$ ,  $J$  necessarily would be greater than one since  $\sigma[r] < \sigma[\rho_c^*]$ . All corrections to the correlation coefficient would consequently affect statistical power negatively. If, however,  $\rho_c > \rho$ , the previous argument no longer applies.

The behaviour of the ratio  $(\rho^*/\sigma[\rho^*]) / (\rho_c^*/\sigma[\rho_c^*])$  under different relevant conditions was consequently examined analytically when  $\rho_c$  assumed under  $H_a$  is estimated via the appropriate correction formula from  $\rho$  assumed under  $H_a$ . In cases where this failed to provide an unambiguous indication of the effect of the particular correction on statistical power, the behaviour of  $J = (\rho^*/\sigma[\rho^*]) / (\rho_c^*/\sigma[\rho_c^*])$  under different values of the relevant parameters was simulated and examined through SAS generated three-dimensional scatter plots of  $J$  against those parameters. These plots would indicate whether corrections to the correlation coefficient affect  $(1-\beta)$ , how they affect  $(1-\beta)$  and under what conditions.

#### 5.4.1 Effect Of Corrections For Attenuation To The Pearson Correlation Coefficient On The A Priori Probabilities $[1-\beta]$ Or Power Of The Tests Of $H_0$ .

##### 5.4.1.1 The Reliability Coefficient Given A Priori By Theoretical Assumption Or Previously Accepted Knowledge

The asymptotic standard error for the partially disattenuated correlation coefficient derived by Bobko and Rieck [1980] for Case III conditions is used to analytically argue as follow:

$$J = (\rho^{\wedge} / \sigma[\rho^{\wedge}]) / (\rho^{\wedge} c / \sigma[\rho^{\wedge} c])$$

$$= (\rho^{\wedge} / \sigma[\rho^{\wedge}]) (\sigma[\rho^{\wedge} c] / \rho^{\wedge} c) \text{-----} 5.3.1$$

Substituting the expression for the partially attenuated validity coefficient [Equation 3.2] in Equation 5.3.1:

$$J = (\rho^{\wedge} / \sigma[\rho^{\wedge}]) (\sigma[\rho^{\wedge} c]) / (\rho^{\wedge} / \sqrt{\rho_{tt} Y})$$

$$= (\rho^{\wedge} / \sigma[\rho^{\wedge}]) (\sigma[\rho^{\wedge} c]) (\sqrt{\rho_{tt} Y} / \rho^{\wedge}) \text{-----} 5.3.2$$

Substituting the expression for the standard error of the uncorrected correlation coefficient [Equation 4.13] in Equation 5.3.2:

$$J = (\rho^{\wedge} / (1 - \rho^{\wedge 2}) / \sqrt{n}) (\sigma[\rho^{\wedge} c]) (\sqrt{\rho_{tt} Y} / \rho^{\wedge})$$

$$= (\rho^{\wedge}) (\sqrt{n} / (1 - \rho^{\wedge 2})) (\sigma[\rho^{\wedge} c]) (\sqrt{\rho_{tt} Y} / \rho^{\wedge}) \text{-----} 5.3.3$$

Substituting Bobko and Rieck's Case III expression for the standard error of the corrected correlation coefficient [Equation 4.25] in Equation 5.3.3:

$$J = (\rho^{\wedge}) (\sqrt{n} / (1 - \rho^{\wedge 2})) (1 / \sqrt{n}) (1 / \sqrt{\rho_{tt} Y}) (1 - \rho^{\wedge 2}) (\sqrt{\rho_{tt} Y} / \rho^{\wedge}) \text{-----} 5.3.4$$

Therefore:

$$J = (\rho^{\wedge} / \sigma[\rho^{\wedge}]) / (\rho^{\wedge} c / \sigma[\rho^{\wedge} c]) = 1 \text{-----} 5.3$$

Equation 5.3 thus leads to the result that the power of the statistical significance test will, under the conditions assumed, remain unaffected by the correction for attenuation for the criterion unreliability only.

### 5.4.1.2 The Reliability Coefficient Obtained From An Independent Sample

The asymptotic standard error for the partially disattenuated correlation coefficient derived by Bobko and Rieck [1980] for Case II conditions is used to analytically argue as follow:

$$\begin{aligned} J &= (\rho^{\wedge}/\sigma[\rho^{\wedge}]) / (\rho^{\wedge}_c/\sigma[\rho^{\wedge}_c]) \\ &= (\rho^{\wedge}/\sigma[\rho^{\wedge}]) (\sigma[\rho^{\wedge}_c]/\rho^{\wedge}_c) \end{aligned} \quad 5.4.1$$

Substituting the expression for the partially attenuated validity coefficient [Equation 3.2] in Equation 5.4.1 and canceling out the term  $\rho^{\wedge}$ :

$$\begin{aligned} J &= (\rho^{\wedge}/\sigma[\rho^{\wedge}]) (\sigma[\rho^{\wedge}_c]) / (\rho^{\wedge}/\sqrt{\rho^{\wedge}_{tt}Y}) \\ &= (\rho^{\wedge}/\sigma[\rho^{\wedge}]) (\sigma[\rho^{\wedge}_c]) (\sqrt{\rho^{\wedge}_{tt}Y}/\rho^{\wedge}) \\ &= (1/\sigma[\rho^{\wedge}]) (\sqrt{\rho^{\wedge}_{tt}Y}) (\sigma[\rho^{\wedge}_c]) \end{aligned} \quad 5.4.2$$

Substituting the expression for the standard error of the uncorrected correlation coefficient [Equation 4.13] in Equation 5.4.2:

$$\begin{aligned} J &= (1/(1-\rho^{\wedge 2})/\sqrt{n}) (\sqrt{\rho^{\wedge}_{tt}Y}) (\sigma[\rho^{\wedge}_c]) \\ &= (\sqrt{n}/(1-\rho^{\wedge 2}) (\sqrt{\rho^{\wedge}_{tt}Y}) (\sigma[\rho^{\wedge}_c]) \end{aligned} \quad 5.4.3$$

Substituting Bobko and Rieck's Case II expression for the asymptotic standard error for the partially disattenuated correlation coefficient [Equation 4.24] in Equation 5.4.3:

$$\begin{aligned} J &= (\sqrt{n}/(1-\rho^{\wedge 2}) (\sqrt{\rho^{\wedge}_{tt}Y}) \{ (1/n) (1/\rho^{\wedge}_{tt}Y) (1-\rho^{\wedge 2})^2 + \\ &\quad (1/4) (1/n) (1/\rho^{\wedge}_{tt}Y^3) (\rho^{\wedge 2}) (1-\rho^{\wedge}_{tt}Y^2)^2 \}^{1/2} \end{aligned} \quad 5.4.4$$

Squaring Equation 5.4.4 and canceling out terms:

$$\begin{aligned} J^2 &= (n/(1-\rho^{\wedge 2})^2 (\rho^{\wedge}_{tt}Y) \{ (1/n) (1/\rho^{\wedge}_{tt}Y) (1-\rho^{\wedge 2})^2 + (1/4) (1/n) (1/\rho^{\wedge}_{tt}Y^3) (\rho^{\wedge 2}) (1-\rho^{\wedge}_{tt}Y^2)^2 \} \\ &= \{ 1 + (1/4) (1/\rho^{\wedge}_{tt}Y^2) (\rho^{\wedge 2}) (1-\rho^{\wedge}_{tt}Y^2)^2 / (1-\rho^{\wedge 2})^2 \} \\ &= \{ 1 + (1/4) (\rho^{\wedge 2} / (1-\rho^{\wedge 2})^2) ((1-\rho^{\wedge}_{tt}Y^2)^2 / \rho^{\wedge}_{tt}Y^2) \} \\ &= \{ 1 + (1/4) (\rho^{\wedge} / (1-\rho^{\wedge 2}))^2 ((1-\rho^{\wedge}_{tt}Y^2) / \rho^{\wedge}_{tt}Y^2) \} \end{aligned} \quad 5.4.5$$

Therefore:

$$J^2 = (\rho^{\wedge 2} / \sigma[\rho^{\wedge}])^2 / (\rho^{\wedge}_c / \sigma[\rho^{\wedge}_c])^2 \geq 1 \quad 5.4.6$$

Since in Equation 5.4:

$$(1/4)(\rho^{\wedge}/(1-\rho^{\wedge 2}))^2((1-\rho^{\wedge} r_{YY}^2)/\rho^{\wedge} r_{YY}^2)^2 \text{ must be } > 0$$

Therefore:

$$J = (\rho^{\wedge}/\sigma[\rho^{\wedge}]) / (\rho^{\wedge} c / \sigma[\rho^{\wedge} c]) \geq 1 \text{-----} 5.4$$

It thus follows that power is negatively affected by Case II corrections for criterion unreliability since Equation 5.4 implies that the critical cutoff under  $H_0$  would translate to a less extreme standard score under  $H_a$  for the corrected correlation coefficient than for the uncorrected coefficient.

#### 5.4.1.3 The Reliability Coefficient Obtained From The Same Sample As $\rho^{\wedge}[X,Y]$

The asymptotic standard error for the partially disattenuated correlation coefficient derived by Bobko and Rieck [1980] for Case I conditions is used to analytically argue as follow:

$$\begin{aligned} J &= (\rho^{\wedge}/\sigma[\rho^{\wedge}]) / (\rho^{\wedge} c / \sigma[\rho^{\wedge} c]) \\ &= (\rho^{\wedge}/\sigma[\rho^{\wedge}]) (\sigma[\rho^{\wedge} c] / \rho^{\wedge} c) \text{-----} 5.5.1 \end{aligned}$$

Substituting Equation 5.4.3 in Equation 5.5.1:

$$J = (\sqrt{n}/(1-\rho^{\wedge 2})) (\sqrt{\rho^{\wedge} r_{YY}^2}) (\sigma[\rho^{\wedge} c]) \text{-----} 5.5.2$$

Squaring Equation 5.5.2:

$$J^2 = (n/(1-\rho^{\wedge 2})^2) (\rho^{\wedge} r_{YY}^2) (\sigma^2[\rho^{\wedge} c]) \text{-----} 5.5.3$$

Substituting Bobko and Rieck's Case I expression for the asymptotic standard error for the partially disattenuated correlation coefficient [Equation 4.23] in Equation 5.4.3:

$$\begin{aligned} J^2 &= \{1 + (1/4)(\rho^{\wedge}/(1-\rho^{\wedge 2}))^2((1-\rho^{\wedge} r_{YY}^2)/\rho^{\wedge} r_{YY}^2)^2\} + \\ &\quad (n/(1-\rho^{\wedge 2})^2) (\rho^{\wedge} r_{YY}^2) \{ (1/n)(1/\rho^{\wedge} r_{YY}^2) (\rho^{\wedge}) [(\rho^{\wedge} r_{YY}^2)(1-2\rho^{\wedge 2}) - (.5)(\rho^{\wedge 2})(1-2\rho^{\wedge 2}-\rho^{\wedge} r_{YY}^2)] \} \\ &= \{1 + (1/4)(\rho^{\wedge}/(1-\rho^{\wedge 2}))^2((1-\rho^{\wedge} r_{YY}^2)/\rho^{\wedge} r_{YY}^2)^2\} + \\ &\quad (\rho^{\wedge}/(1-\rho^{\wedge 2})^2) (1/\rho^{\wedge} r_{YY}^2) [(\rho^{\wedge} r_{YY}^2)(1-2\rho^{\wedge 2}) - (.5)(\rho^{\wedge 2})(1-2\rho^{\wedge 2}-\rho^{\wedge} r_{YY}^2)] \} \text{-----} 5.5 \end{aligned}$$

Equation 5.5, although it could be simplified further, fails to provide an unambiguous indication of the effect of the [Case I] correction for criterion unreliability on statistical power. The reaction of  $J = (\rho^2/\sigma[\rho^2]) / (\rho_c^2/\sigma[\rho_c^2])$  to changes in  $\rho[X,Y]$ ,  $n$  and  $\rho_{tt}Y$  was subsequently calculated and examined through SAS generated three-dimensional scatter plots of  $J$  against those parameters. Figure 5.36 and Figure 5.37 suggest Case 1 partial corrections for attenuation to negatively affect statistical power for all permissible combinations of  $\rho[X,Y]$  and  $\rho_{tt}Y$  and values of  $n$ .  $J$  is consistently greater than 1 which implies that the critical cutoff under  $H_0$  translates to a less extreme standard score under  $H_a$  for the corrected correlation coefficient than for the uncorrected coefficient. Figures 5.36 and 5.37 furthermore reveal that the negative effect of Case I partial corrections for attenuation on statistical power increases curvilinearly with increases in the effect size assumed for the uncorrected correlation coefficient over all levels of criterion reliability.

#### 5.4.2 Effect Of Corrections For Restriction Of Range To The Pearson Correlation Coefficient On The A Priori Probabilities $[1-\beta]$ Or Power Of The Tests Of $H_0$ .

##### 5.4.2.1 Case 2 [Case A] Restriction Of Range

The asymptotic standard error for the correlation coefficient corrected for Case 2 [Case A] restriction of range, derived by Bobko and Rieck [1980], is used to analytically argue as follow:

$$J = (\rho^2/\sigma[\rho^2]) / (\rho_c^2/\sigma[\rho_c^2]) \\ = (\rho^2/\sigma[\rho^2])(\sigma[\rho_c^2]/\rho_c^2) \text{-----} 5.6.1$$

Substituting the expression for the asymptotic standard error for the correlation coefficient corrected for Case 2 [Case A] restriction of range, derived by Bobko and Rieck [1980] [Equation 4.35]

$$J = (\rho^2/\sigma[\rho^2])(K/(1+\rho^2K^2-\rho^2)^{3/2})(\sigma[\rho^2])(1/\rho_c^2) \text{-----} 5.6.2$$

Substituting the correction formula for Case 2 [Case A] restriction of range [Equation 3.4] in Equation 5.6.2:

$$J = (\rho^2)(K/(1+\rho^2K^2-\rho^2)^{3/2})((1+\rho^2K^2-\rho^2)^{1/2}/\rho^2K) \\ = (1+\rho^2K^2-\rho^2)^{-1} \text{-----} 5.6$$

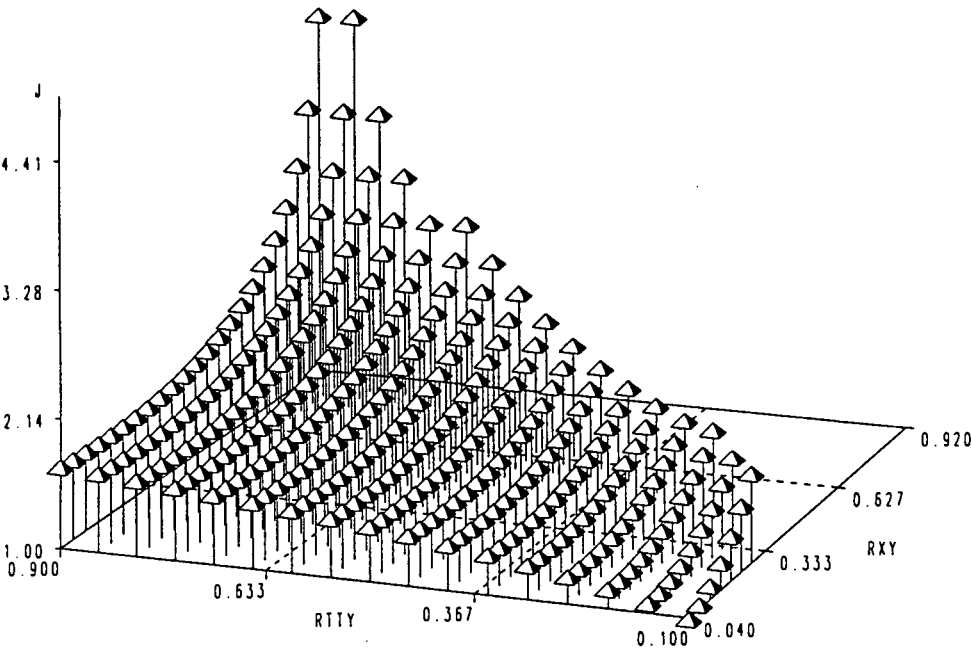


Figure 5.36:  $J = (\rho^{\wedge 2}/\sigma[\rho^{\wedge}]) / (\rho^{\wedge}_c/\sigma[\rho^{\wedge}_c])$  as a function of  $\rho^{\wedge}[X,Y]$  and  $\rho^{\wedge}_{tt}Y$  for  $n$  fixed at 10

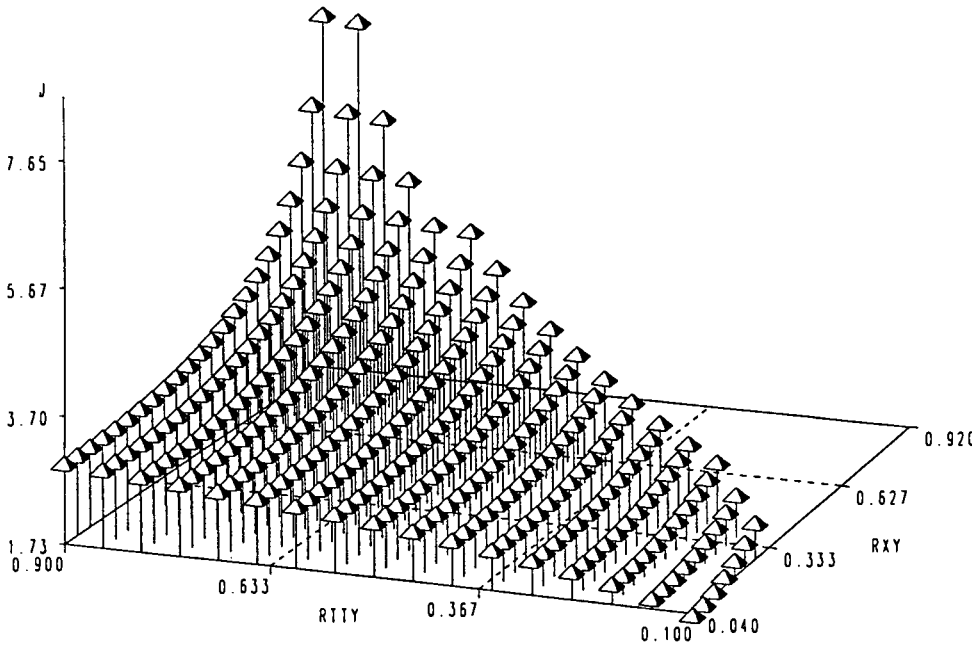


Figure 5.37:  $J = (\rho^{\wedge 2}/\sigma[\rho^{\wedge}]) / (\rho^{\wedge}_c/\sigma[\rho^{\wedge}_c])$  as a function of  $\rho^{\wedge}[X,Y]$  and  $\rho^{\wedge}_{tt}Y$  for  $n$  fixed at 90



Therefore it follows that:

if  $K = \sigma[X]/\sigma[x] = 1$  then  $(1 + \rho^2 K^2 - \rho^2)^{-1} = 1$  for all  $\rho$

and

if  $K = \sigma[X]/\sigma[x] > 1$  then  $(1 + \rho^2 K^2 - \rho^2)^{-1} < 1$  for all  $\rho$

It thus follows that the power of the statistical significance test would increase, should the correction for Case 2 [Case A] restriction of range be applied to the observed, restricted correlation coefficient and the uncorrected effect size estimate.

#### 5.4.2.2 Case C Restriction Of Range

The asymptotic standard error for the correlation coefficient corrected for Case C restriction of range, derived by Allen and Dunbar [1990], is sufficiently complex to eliminate, for all practical purposes, any possibility of finding an analytical solution to the question in hand. Values of  $J$  were thus calculated for changes in  $\rho^{[x,z]}$ ,  $\rho^{[y,z]}$ ,  $n$  and the initial uncorrected effect size estimate  $\rho^{[x,y]}$ . Figures 5.38 - 5.43 provide a pictorial representation of the reaction of  $J$  to changes in  $\rho^{[x,z]}$ ,  $\rho^{[y,z]}$ ,  $n$  and  $\rho^{[x,y]}$ . Figures 5.38 - 5.43 suggest that Case C corrections for restriction of range can affect the statistical power of significance tests both positively and negatively depending on the applicable parameter settings. Corrections for Case C restriction of range [fortunately] seem to have maximum adverse impact on statistical power when  $\rho^{[x,y]}$  is high and  $\rho^{[x,z]}$  and  $\rho^{[y,z]}$  is low.  $K$  seems to affect power by aggravating the effect of the aforementioned parameters as  $K$  increases [i.e. as the selection ratio decreases]. In contrast, corrections for Case C restriction of range seem to have maximum beneficial impact on statistical power under those conditions where Case C restriction of range corrections have their greatest impact, namely when  $\rho^{[x,y]}$  is low and  $\rho^{[x,z]}$  and  $\rho^{[y,z]}$  is high.  $K$  again seems to affect power by aggravating the effect of the aforementioned parameters as  $K$  increases [i.e. as the selection ratio decreases].

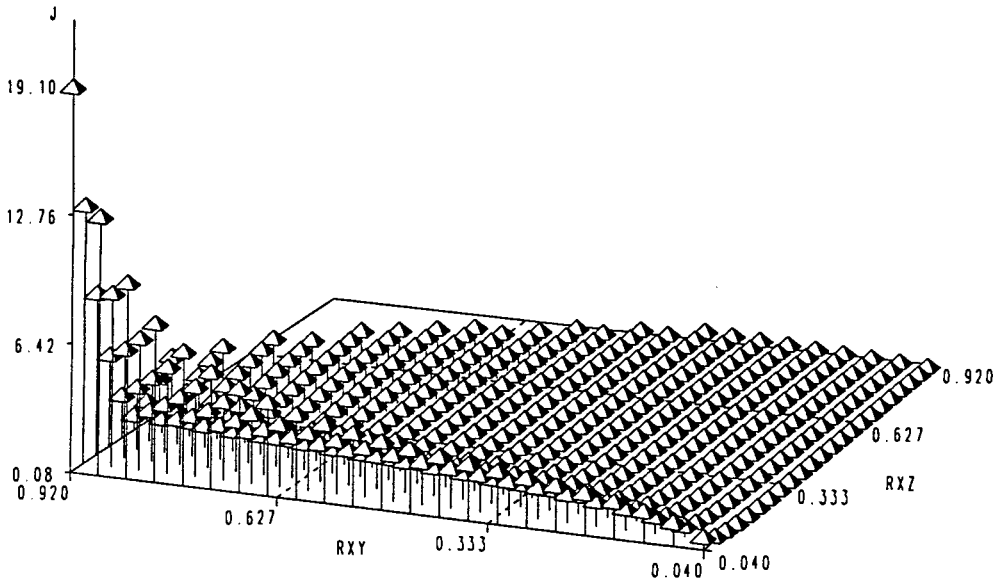


Figure 5.38:  $J = (\hat{\rho}/\sigma[\hat{\rho}]) / (\hat{\rho}_c/\sigma[\hat{\rho}_c])$  under Case 3[i] [Case C[i]] restriction of range as a function of  $\hat{\rho}[x,z]$  and  $\hat{\rho}[x,y]$  for  $n$  fixed at 10,  $K$  fixed at 4 and  $\hat{\rho}[y,z]$  fixed at 0.25

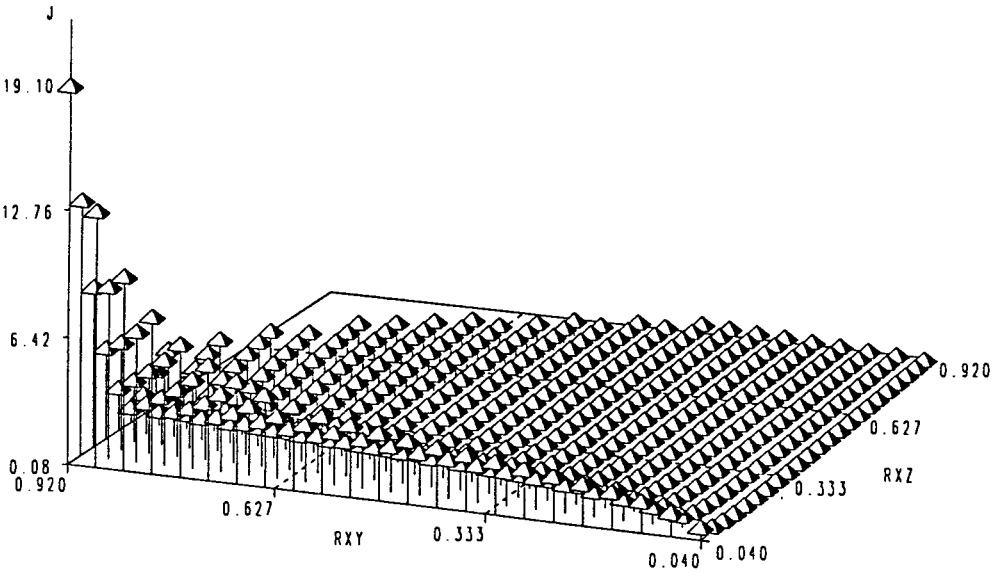


Figure 5.39:  $J = (\hat{\rho}/\sigma[\hat{\rho}]) / (\hat{\rho}_c/\sigma[\hat{\rho}_c])$  under Case 3[i] [Case C[i]] restriction of range as a function of  $\hat{\rho}[x,z]$  and  $\hat{\rho}[x,y]$  for  $n$  fixed at 90,  $K$  fixed at 4 and  $\hat{\rho}[y,z]$  fixed at 0.25

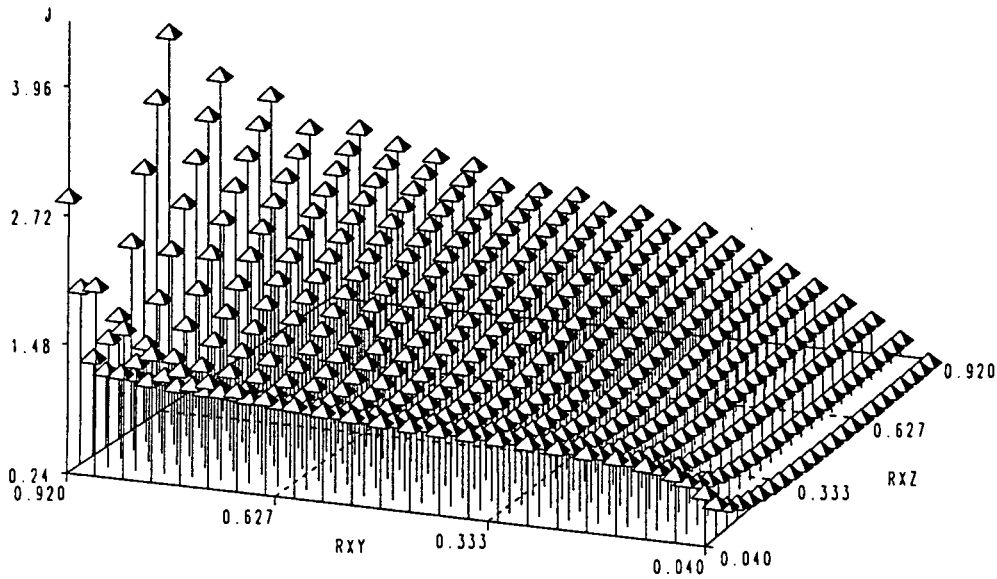


Figure 5.40:  $J = (\rho^{\hat{}}/\sigma[\rho^{\hat{}}]) / (\rho^{\hat{}}_c/\sigma[\rho^{\hat{}}_c])$  under Case 3[i] [Case C[i]] restriction of range as a function of  $\rho^{\hat{}}[x,z]$  and  $\rho^{\hat{}}[x,y]$  for  $n$  fixed at 90,  $K$  fixed at 2 and  $\rho^{\hat{}}[y,z]$  fixed at 0.15

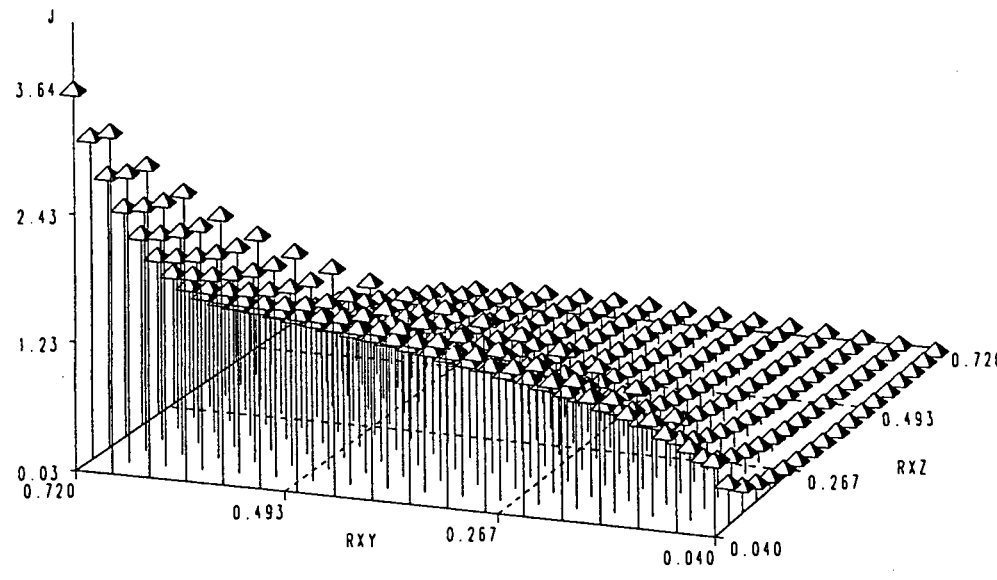


Figure 5.41:  $J = (\rho^{\hat{}}/\sigma[\rho^{\hat{}}]) / (\rho^{\hat{}}_c/\sigma[\rho^{\hat{}}_c])$  under Case 3[i] [Case C[i]] restriction of range as a function of  $\rho^{\hat{}}[x,z]$  and  $\rho^{\hat{}}[x,y]$  for  $n$  fixed at 90,  $K$  fixed at 2 and  $\rho^{\hat{}}[y,z]$  fixed at 0.65

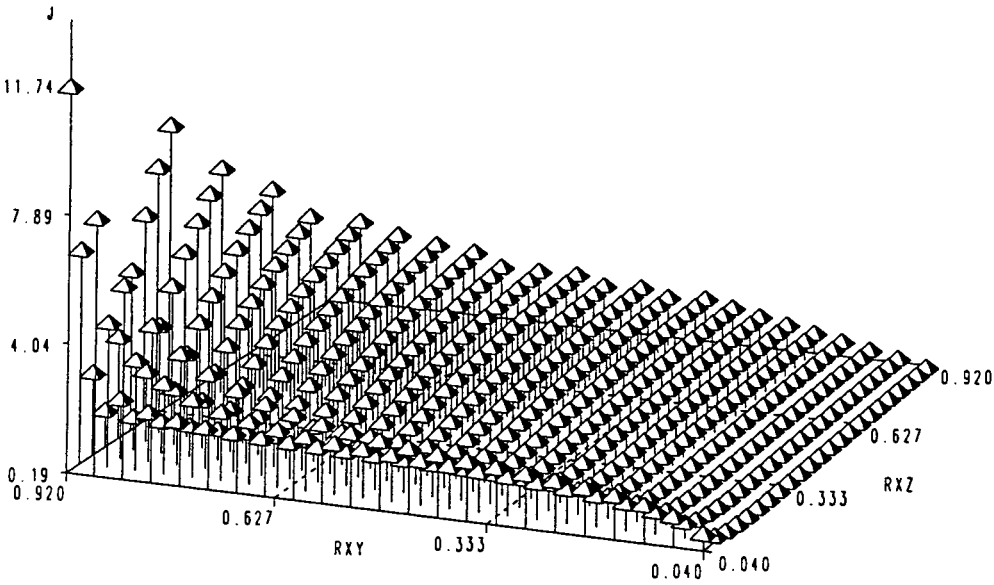


Figure 5.42:  $J = (\rho^{\wedge}/\sigma[\rho^{\wedge}]) / (\rho^{\wedge}_c/\sigma[\rho^{\wedge}_c])$  under Case 3[i] [Case C[i]] restriction of range as a function of  $\rho^{\wedge}[x,z]$  and  $\rho^{\wedge}[x,y]$  for  $n$  fixed at 90,  $K$  fixed at 4 and  $\rho^{\wedge}[y,z]$  fixed at 0.15

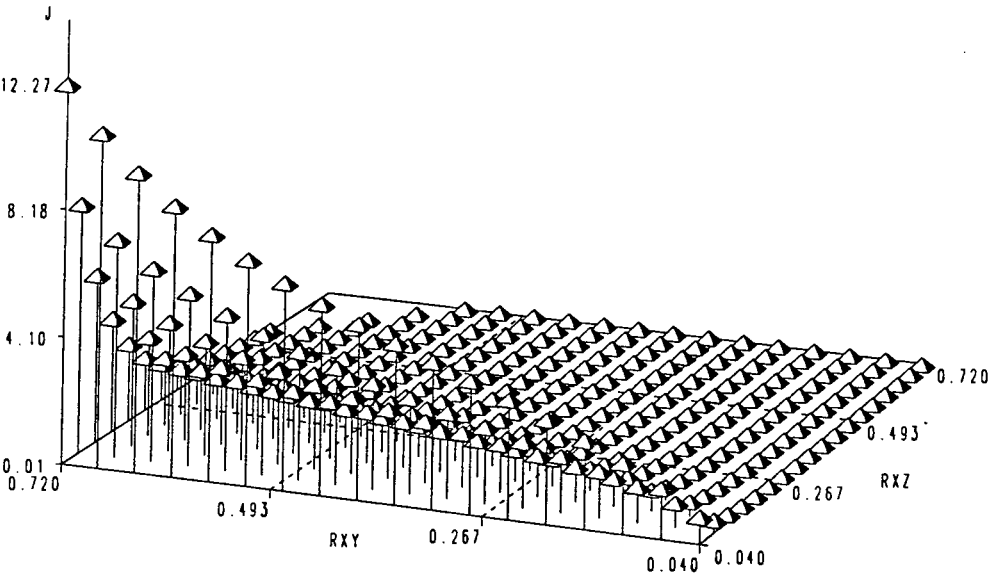


Figure 5.43:  $J = (\rho^{\wedge}/\sigma[\rho^{\wedge}]) / (\rho^{\wedge}_c/\sigma[\rho^{\wedge}_c])$  under Case 3[i] [Case C[i]] restriction of range as a function of  $\rho^{\wedge}[x,z]$  and  $\rho^{\wedge}[x,y]$  for  $n$  fixed at 90,  $K$  fixed at 4 and  $\rho^{\wedge}[y,z]$  fixed at 0.65

### 5.4.3 Effect Of Double Corrections For Attenuation And Restriction Of Range To The Pearson Correlation Coefficient On The A Priori Probabilities $[1-\beta]$ Or Power Of The Tests Of $H_0$ .

Substituting the asymptotic standard error for the correlation coefficient double corrected for [Case 2 [Case A]] restriction of range and criterion unreliability derived by Bobko [1983] in  $J$  and analytically simplifying the ratio would unlikely yield any definitive conclusion on the reaction of  $J$  to changes in the initial uncorrected effect size estimate,  $n$ ,  $\rho_{ttY}$  and  $K$ . The computational solution of  $J$  to changes in the relevant parameters combined with pertinent three dimensional scatter plots once more seems to be the more productive option. Figures 5.44 - 5.47 depict the reaction of  $J$  to changes in the uncorrected ES estimate,  $\rho_{ttY}$ ,  $n$  and  $K$ . Figures 5.44 - 5.47 reveal that the probability of the corrected correlation coefficient exceeding the critical cut off correlation, conditional on a specific value for  $\rho$  assumed under  $H_a$ , is smaller than the corresponding conditional probability for the double corrected correlation coefficient for all  $\rho^*[x,y]$ ,  $\rho_{ttY}$  and for  $K > 1$ . Since  $J < 1$  for all  $\rho^*[x,y]$ ,  $\rho_{ttY}$  and for  $K > 1$ , statistical power is improved by the double correction for restriction of range and criterion unreliability, provided the uncorrected ES estimate is also corrected via the appropriate correction formula [Equation 3.10 or Equation 3.13] to obtain the corrected ES estimate. Maximum power benefits are obtained by the double correction when  $\rho^*[x,y]$  is high [i.e. the initial ES estimate is high],  $\rho_{ttY}$  is low and  $K$  is high [i.e. the selection ratio is small]. Although  $n$  affects power, it does not seem to affect the change in power brought about by the double correction to the correlation coefficient

## 5.5 SUMMARY

Chapter 5 presented an introductory synopsis of the basic logic underlying classical/Fisherian statistical hypothesis/significance testing.

The consequences of corrections for attenuation and/or restriction of range to the Pearson correlation coefficient on decisions on the validity of statistical null hypotheses were subsequently examined. Two quantities were identified to have particular relevance, namely the empirically derived exceedence probabilities and the power of the tests of  $H_0$ . The quantities  $G = (\rho^*_c/\sigma[\rho^*_c])/(\rho^*/\sigma[\rho^*])$  and  $J = (\rho_t/\sigma[\rho^*])/(\rho_c/\sigma[\rho^*_c])$  were defined and shown to constitute appropriate indicators of the effect of the corrections on  $\alpha_b$  and  $1-\beta$ . The reaction of the ratios  $G$  and  $J$  to corrections to the validity coefficient were examined analytically where possible. However, in cases where this proved to be a too formidable task, the behaviour of  $G$  and  $J$  under different values of the relevant parameters was examined through computer generated plots of  $G$  and  $J$  against those parameters.

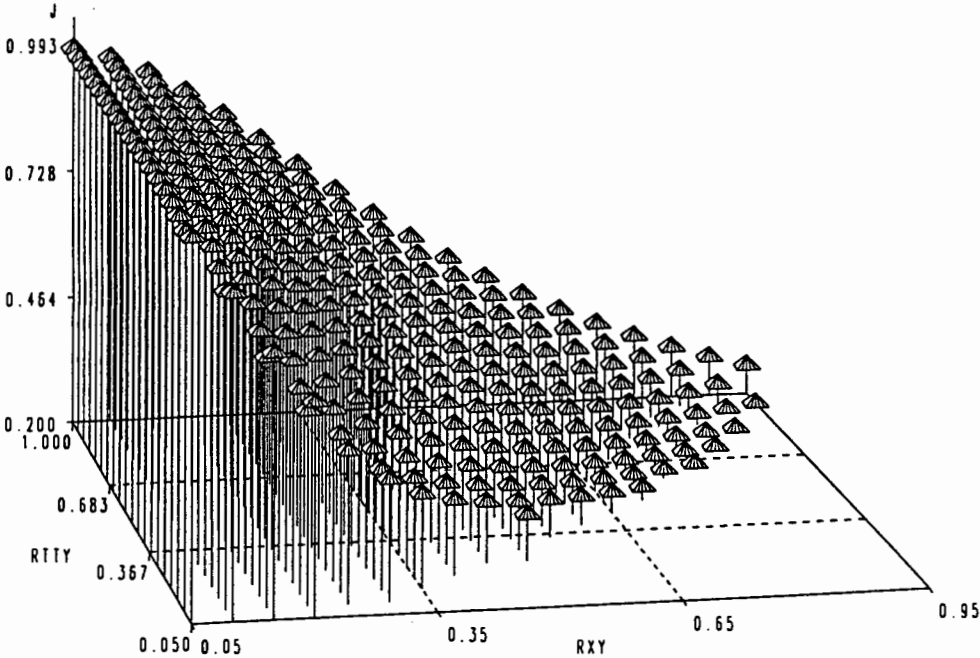


Figure 5.44:  $J = (\hat{\rho}/\sigma[\hat{\rho}]) / (\hat{\rho}_c/\sigma[\hat{\rho}_c])$  under the joint correction for Case 2 [Case A] restriction of range and criterion unreliability as a function of the initial uncorrected effect size estimate and  $\rho_{\text{TTY}}$  for  $n$  fixed at 10 and  $K$  fixed at 2

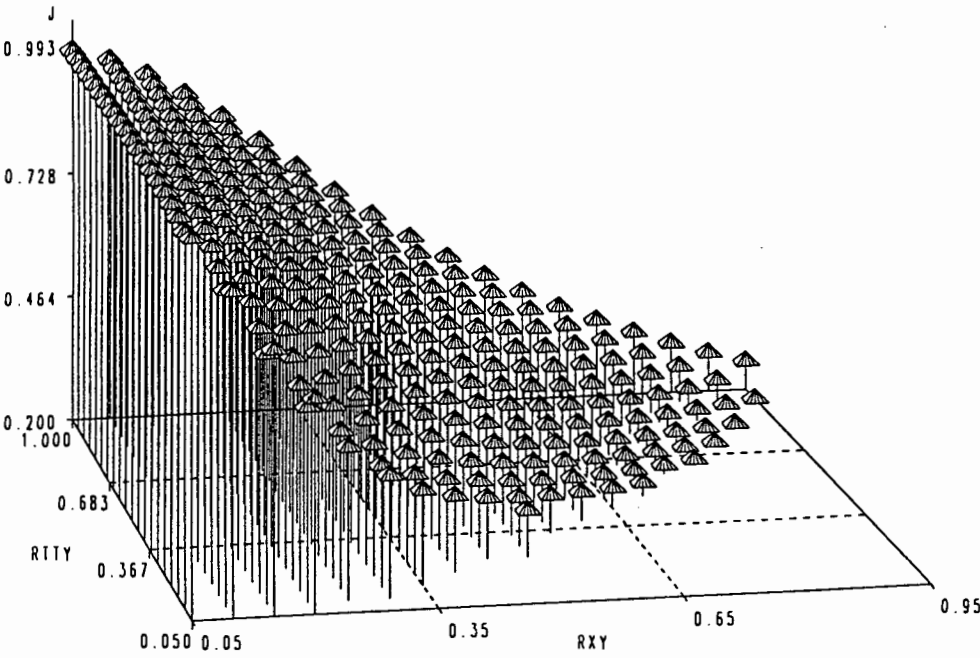


Figure 5.45:  $J = (\hat{\rho}/\sigma[\hat{\rho}]) / (\hat{\rho}_c/\sigma[\hat{\rho}_c])$  under the joint correction for Case 2 [Case A] restriction of range and criterion unreliability as a function of the initial uncorrected effect size estimate and  $\rho_{\text{TTY}}$  for  $n$  fixed at 90 and  $K$  fixed at 2



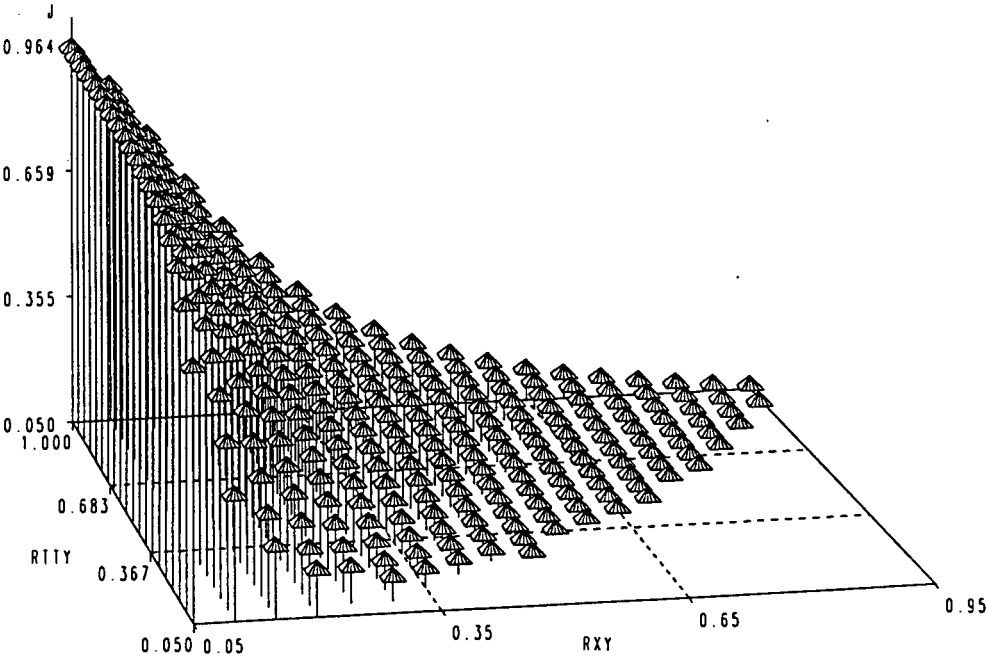


Figure 5.46:  $J = (\hat{\rho} / \sigma[\hat{\rho}]) / (\hat{\rho}_c / \sigma[\hat{\rho}_c])$  under the joint correction for Case 2 [Case A] restriction of range and criterion unreliability as a function of the initial uncorrected effect size estimate and  $\rho_{\text{TTY}}$  for  $n$  fixed at 10 and  $K$  fixed at 4

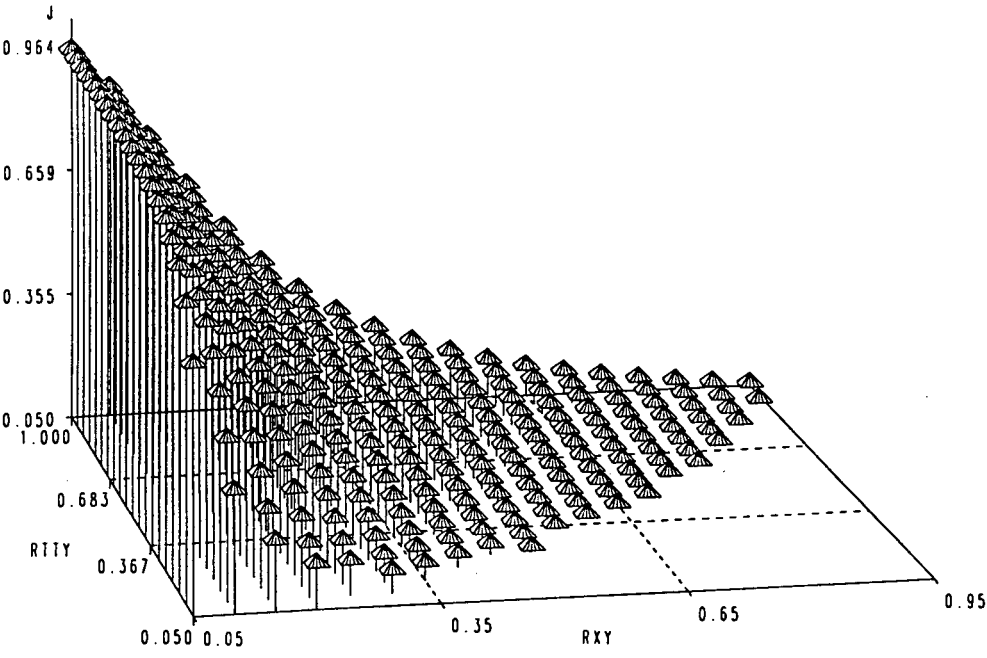


Figure 5.47:  $J = (\hat{\rho} / \sigma[\hat{\rho}]) / (\hat{\rho}_c / \sigma[\hat{\rho}_c])$  under the joint correction for Case 2 [Case A] restriction of range and criterion unreliability as a function of the initial uncorrected effect size estimate and  $\rho_{\text{TTY}}$  for  $n$  fixed at 90 and  $K$  fixed at 4

## CHAPTER 6

### EFFECT OF STATISTICAL CORRECTIONS TO THE VALIDITY COEFFICIENT ON THE PARAMETERS OF THE DECISION FUNCTION

The purpose of the following discussion is to extend the logic underlying the corrections for restriction of range and/or criterion unreliability applied to the validity coefficient to the decision rule in terms of which applicants are screened for employment and to examine the consequences of applying such correction for actual selection decision making. A synopsis of the logic underlying selection decision making will, however, first be presented so as to establish the necessary theoretical foundation for the subsequent analysis.

#### 6.1 THE DECISION FUNCTION IN PERSONNEL SELECTION

Personnel selection is necessitated by the combined effect of inter-individual differences amongst applicants on those attributes that would determine their eventual job performance and the selecting organisation's desire to maximise performance. The desire to maximise performance implies work success as the ultimate/final institutional criterion in terms of which applicants for employment should ideally be evaluated and on which they should ideally be compared so as to arrive at an institutionally rational selection decision. Personnel selection is, however, complicated by the obvious fact that information on the ultimate institutional criterion can never be available at the time of the selection decision. The only solution to this dilemma, apart from reducing selection to random assignment, is to base the decision on relevant substitute information that is assessable prior to the selection decision. Even if no direct information on the criterion ever enters the selection decision making process, the criterion nonetheless always remains the focus of interest in selection assessment. This seemingly innocent and too often forgotten fact, moreover, has significant implications for the interpretation and evaluation of information entering the selection decision.

Only two basic options exist in terms of which such substitute information can be generated. In terms of the first option substitute information would be obtained through low or high fidelity simulations of the job content [or key performance areas] developed from systematically constructed job descriptions. The first option thus requires no understanding of the reasons underlying differences in performance. It does, however, require an understanding of the demands that collectively constitute the job. It in essence argues that if the simulation would succeed in creating conditions that place similar demands on people as the job would [although those conditions need not look similar to the job content], measurements of performance in the simulation should be systematically related to eventual job

performance. Although the simulation should, under the aforementioned assumptions, logically invoke the same attributes [or competencies] as would be required to succeed in the job, those critical attributes and their interaction need not necessarily be understood. The second option, in contrast, requires that the same job description should be examined again, but now with the explicit purpose of inferring those qualities of job incumbents that are required to succeed in the job. The second option thus requires a comprehensive understanding of the reasons underlying performance in addition to clarity on the fundamental nature of the key performance areas comprising the job. It in essence argues that if the hypotheses on job performance derived from the job description would be valid [i.e. if the attributes suggested by the job description to be critical for success, are in fact competencies], measurements of these critical characteristics [if combined in accordance with the dynamics of the performance theory] should be systematically related to eventual job performance.

Information obtained via either one of the aforementioned two options derives its ability to act as a substitute and thereby its relevance through the extent to which it correlates with a valid representation or operationalisation of the ultimate criterion [i.e. to the extent that the substitute statistically explains variance in the criterion]. This has the important [but seemingly too easily forgotten] implication that any measurement procedure [interpreted in the widest possible meaning of the term] which delivers [content/construct valid] measurements of any knowledge domain, skill, ability, personality trait, motive or style, but which bases its claim to relevance solely on logic, will remain on the level of conjecture until empirically proven otherwise. Stated differently, attributes/characteristics do not qualify as competencies simply because they are designated as such. Competencies are underlying characteristics of people [or behaviour] that have been *empirically shown* to be [causally] related to effective job performance [Boyatzis, 1982; Spangenberg, 1990].

Relevance thus implies a systematic relationship between the criterion and the substitute information. The identification of relevant substitute information therefore, by definition, creates the possibility of estimating the expected criterion performance and/or probability of success/failure conditional on the information content. One should in fact argue that the foregoing interpretation of the term relevance flows logically from the need to predict the criterion on which the selection decision ideally should be based, but which is not directly assessable at the time of the decision. A necessary prerequisite, however, to achieve such criterion referenced interpretation of the substitute information, is that the nature of the relationship between the substitute information and the criterion be known.

Effective [although not necessarily efficient] personnel selection is therefore possible if, and only if:

- [substitute] information is available at the time of the selection decision that is systematically related to the ultimate/final criterion of work success [i.e. relevant information]; and



- the nature of the relationship is at least subjectively/clinically but preferably statistically/actuarially] understood.

The relevance of substitute information is established through an extensive validation study as a form of applied explanatory research [Ellis & Blustein, 1991; Landy, 1986; Schmitt & Landy, 1993]. The credibility of the any claims to relevance resulting from a validation study will consequently in the final analysis depend on the unassailableness, scientific rationality or validity [Cook, Campbell & Perrachio, 1991] of the research methodology through which hypotheses derived from a job description were investigated.

Once the case for the relevance of the substitute information has been successfully argued, the question on how to combine such information to arrive at a decision arises [strictly speaking, however, these two issues can not be completely separated]. Two, and only two, basic option exist in terms of which information can be combined for decision making. Both options require that the nature of the relationship between the criterion and the substitute information be understood. The two options, however, differ in the way they express their understanding of the criterion - information relationship. The first option could be termed a judgmental, subjective or clinical mode of information combination since the decision outcome is derived from human judgement based on an inexplicit and unstandardised decision rule. The second option could be termed a mechanical, statistical or actuarial mode of information combination since an explicit and standardised rule or formula dictates the decision outcome [Gatewood & Feild, 1994]. Within the mechanical option a number of different actuarial selection strategies can be distinguished. A selection strategy in the current context refers to an explicit rule which determines, conditional on obtained information, the assignment of applicants to one of three possible outcomes, namely terminal rejection, acceptance or further investigation [Cronbach & Gleser, 1965; Gatewood & Feild, 1994]. All actuarial [i.e. prescriptive] selection strategies can be reduced to a strategy matrix [Cronbach & Gleser, 1965] expressing its allocation of applicants to treatments in terms of probabilities, restricted to either 0 or 1, conditional on obtained information. The nature of a selection strategy and thus the structure of the strategy matrix depends on:

- whether compensation will be permitted in the combination of information; and
- whether multi-stage sequential assessment will be permitted or required.

The first consideration impacts on the number of ways [Coxon, 1982; Schiffman, Reynolds & Young, 1981] the strategy matrix contains. If an information x decision format is assumed for the most basic strategy matrix, such a matrix would comprise two ways. If compensation is allowed, the strategy matrix will always comprise two ways. The information way can be presented in terms of k class intervals or in terms of a continuum. If compensation in the combination of information is disallowed, the number of ways will increase from two to the number of dimensions information was obtained on. Multi-stage assessment will affect the number of columns in the decision way of the strategy matrix and

the number of matrixes. Sequential assessment will result in an array of information x decision strategy matrixes, each consisting of three columns in the decision way, except in the final matrix.

The content of the strategy matrix [i.e. the distribution of the conditional probabilities of acceptance (either 0 or 1)], in turn will depend on:

- whether quotas are in force;
- the aspiration level of the decision maker;
- the relevance of the information on which the strategy is based;
- the nature of the relationship between the criterion and the selection information;
- the presence of group x information interaction effects; and
- the size of the applicant group;

The following [actuarial] selection strategies are typically distinguished [Gatewood & Feild, 1994; Milkovich & Boudreau, 1994]:

- regression strategy [single stage; compensation allowed]
- multiple cutoff [single stage; limited compensation allowed]
- multiple hurdle [multi-stage; no compensation allowed];
- profile comparison [single stage; compensation depending on comparison technique]
- combinations of the above.

Attention will henceforth solely be focused on the regression strategy. Three basic reasons motivated the decision to restrict the analysis to the regression strategy only. To consider all the aforementioned strategies in the current study was, firstly, considered unnecessarily ambitious. Although all the selection strategies are worthy of consideration, they need not all be examined simultaneously in a single study. The present study, secondly, chose to focus on the regression strategy because it forms the implicit basis of all the other selection strategies and because it constitutes the strategy explicitly assumed in the Brogden-Cronbach-Gleser utility analysis [Boudreau, 1991] and the Cleary-model on selection fairness [Cleary, 1968]. The goal of the multiple regression strategy is to find a weighted linear combination of the individual information sources that minimises the sum of the squared deviations between the linear combination and the actual criterion and thus that maximally correlates with the actual criterion [Cohen & Cohen, 1983; Tabachnick & Fidell, 1989].

A two way information x treatment strategy matrix is thus assumed. The treatment way of the matrix comprises the two possible terminal decisions to either reject or accept the applicant. The information way of the strategy matrix could either consist of  $k$  class intervals formed on the expected criterion performance [i.e.  $E[Y | X_i]$ ] scale or the conditional probability of success [i.e.  $P[Y \geq Y_c | X = X_i]$ ] scale or, alternatively, consist of one of the aforementioned continuous scales.

A critical cutoff score  $Y_C$  exists on the criterion, reflecting the aspiration level of the decision maker. The ensuing argument will be based on the assumptions that the critical criterion cutoff value,  $Y_C$ , is defined on a scale unaffected by [systematic and/or random] measurement error.  $Y_C$  thus reflects the minimum actual contribution considered acceptable. This interpretation of  $Y_C$  has two important consequences. It firstly implies that the criterion is positive in the sense that higher scores represent a more positive evaluation. Consequently all  $Y \geq Y_C$  would be considered successful irrespective of the position of  $Y_C$  in the criterion distribution. It secondly implies that success is not interpreted normatively but rather criterion construct referenced. This assumption implies that the critical cutoff will remain numerically unaffected by the statistical removal of random measurement error from the criterion and/or the statistical reversal of the effect of explicit or implicit selection. Consequently criterion performance after the corrections still would be considered successful only if it exceeds  $Y_C$ , irrespective of the [changed] position of  $Y_C$  in the corrected criterion distribution. A normative interpretation, in contrast, would inevitably result in a numerical change in  $Y_C$ .

Similarly a critical cutoff score  $\alpha_s$  exists on the conditional probability of success scale, also, reflecting the aspiration level of the decision-maker. The critical probability  $\alpha_s$  reflects the minimum chance of success at which the decision-maker still regards it as worthwhile to accept an applicant for employment. The probability  $[1 - \alpha_s]$  reflects the maximum risk the decision-maker is prepared to take when considering applicants for employment.

Two simple decision rules result if the number of applicants that may be permissibly selected is not restricted by the existence of any selection quotas:

- "if  $E[Y | X_i] \geq Y_C$  then accept; else reject"; or
- "if  $P[Y \geq Y_C | X = X_i] \geq \alpha_s$  then accept; else reject".

The second first and second decision rule will result in the acceptance of the same number  $[N_s]$  and the identical applicants if  $\alpha_s = 0.50$ . The second decision rule, however, constitutes a more demanding decision rule than the first to the extent that  $\alpha_s$  is set above 0.50. The selection ratio  $[SR = N_s/N]$  under the second decision rule will therefore always be smaller than it would be under the first selection rule, if  $\alpha_s > 0.50$ . In terms of the strategy matrix the aforementioned two decision rules translate to:

- "if  $E[Y | X_i] \geq Y_C$  then  $P[\text{decision=accept}] = 1$ ; else  $P[\text{decision=accept}] = 0$ "; or
- "if  $P[Y \geq Y_C | X_i] \geq \alpha_s$  then  $P[\text{decision=accept}] = 1$ ; else  $P[\text{decision=accept}] = 0$ ".

The existence of selection quotas restricting the number of applicants that may be permissibly selected, naturally affect the aforementioned decision rules. Selection quotas may be expressed in terms of the number of vacancies that may be filled  $[N_v]$ , in terms of the number of applicants that may be selected  $[N_{ps}]$  or in terms of a permissible selection ratio  $[SR = N_v/N = N_{ps}/N]$ . A number of options exist in terms of which  $N_{ps}$  applicants could be selected from the  $N_s$  applicants which would qualify for unrestricted selection. From an institutional perspective, strict top down selection constitutes the



option that maximises selection utility [and in that sense the most advantageous option] when selection is restricted by quotas [Boudreau, 1991; Gatewood & Feild, 1994]. The following two decision rules result when selection is restricted by quotas and a strict top down strategy is assumed [RO refers to the rank order of the  $N_s$  applicants which would qualify for unrestricted selection, with the highest rank allocated to the applicant with the highest expected criterion score or highest conditional probability]:

- "if  $E[Y | X_i] \geq Y_c$  and  $RO \leq N_v$  then accept; else reject"; or
- "if  $P[Y \geq Y_c | X = X_i] \geq \alpha_s$  and  $RO \leq N_v$  then accept; else reject".

For the purpose of the ensuing exploration of the possible effects of extending the logic underlying statistical corrections to the validity coefficient to the parameters of the decision function, a simple linear regression model will be assumed. The decision to restrict the ensuing analysis to simple linear regression only is entirely motivated by the desire to initially limit the problem to its most basic, and therefore most manageable, form. This decision should, therefore, not be interpreted as an implication that the majority of actual selection decision-making is based on the simple linear regression model. Clearly there exists a need to eventually expand the analysis to multiple regression and other selection strategies. Both  $X$  and  $Y$  can, however, represent composite variables. To some, albeit limited, extent the multi-variate nature of actual selection decision-making is thus thereby acknowledged. The simple linear regression model assumes a linear relationship between the criterion  $Y$  and an information/predictor variable  $X$  that can, as a population model, be expressed as Equation 6.1.

$$Y_i = \alpha + \beta X_i + \varepsilon_i \text{-----} 6.1$$

The residual  $\varepsilon_i$  in the preceding model represents that portion of the observed value  $Y_i$  that can not be explained from  $X_i$  in terms of a linear relationship between  $Y$  and  $X$ . The error term  $\varepsilon_i$  thus accounts for variables other than  $X$  that explain variance in  $Y$  but are not included in the model, random error in  $Y$  attributable to measurement error and the inappropriateness of the linear model [Berenson, Levine & Goldstein, 1983; Younger, 1979]. Since  $\varepsilon_i$  is unknown,  $Y_i$  can not be determined exactly through its linear relation with  $X_i$ . The parameters  $\alpha$  and  $\beta$ , representing the  $Y$  axis intercept and  $X$  axis slope respectively, are determined through Equation 6.2 and 6.3 so as to minimise  $\Sigma \varepsilon^2 = \Sigma [Y_i - E[Y_i | X_i]]$  [Cohen & Cohen, 1983; Tabachnick & Fidell, 1989].

$$\alpha = \mu[Y] - \beta \mu[X] \text{-----} 6.2$$

$$\beta = \rho[X,Y](\sigma[Y]/\sigma[X]) \text{-----} 6.3$$

By substituting Equations 6.2 and 6.3 in Equation 6.1, the latter can be rewritten as Equation 6.4.

$$E[Y | X_i] = \alpha + \beta X_i$$

$$\begin{aligned}
&= \mu[Y] - \beta\mu[X] + \beta X_i \\
&= \mu[Y] + \beta[X_i - \mu[X]] \\
&= \mu[Y] + \rho[X,Y](\sigma[Y]/\sigma[X])(X_i - \mu[X]) \text{-----} 6.4
\end{aligned}$$

Equation 6.4 assumes the information variable  $X$  to be fixed at specific levels whereas, at each fixed value of  $X$ , the criterion variable  $Y$  is assumed to be a random variable. In addition, each of these sub-populations of  $Y$  values conditional on a fixed  $X$ , are assumed to follow a normal distribution with mean  $\mu[Y|X]$  and variance  $\sigma^2[Y|X]$ . Equation 6.4, by definition, assumes the conditional criterion means  $\mu[Y|X]$  to change linearly with changes in  $X$ . Equation 6.4, furthermore, assumes the conditional criterion variance  $\sigma^2[Y|X]$  to remain constant over changes in  $X$  [Cohen & Cohen, 1983; Tabachnick & Fidell, 1989]. The average squared deviation [i.e. variance] about the regression line  $\sigma^2[Y|X]$  is expressed by Equation 6.5.

By definition:

$$\varepsilon = Y_i - E[Y|X_i] \text{-----} 6.5.1$$

Taking the expected value of the squared deviation of the observed criterion scores from the expected criterion score conditional on the observed predictor score and substituting Equation 6.1:

$$\begin{aligned}
E[\varepsilon^2] &= E[Y_i - E[Y|X_i]]^2 \\
&= E[Y_i - (\alpha + \beta X_i)]^2 \text{-----} 6.5.1
\end{aligned}$$

Substituting Equation 6.2 in Equation 6.5.1:

$$\begin{aligned}
E[\varepsilon^2] &= E[Y_i - (\mu[Y] - \beta\mu[X]) - \beta X_i]^2 \\
&= E[Y_i - \mu[Y] + \beta\mu[X] - \beta X_i]^2 \\
&= E[(Y_i - \mu[Y]) - \beta(X_i + \mu[X])]^2 \text{-----} 6.5.2
\end{aligned}$$

Factoring out Equation 6.5.2:

$$\begin{aligned}
E[\varepsilon^2] &= E[(Y_i - \mu[Y])^2 - 2\beta(Y_i - \mu[Y])(X_i + \mu[X]) + \beta^2(X_i + \mu[X])^2] \\
&= E[Y_i - \mu[Y]]^2 - 2\beta E[(Y_i - \mu[Y])(X_i + \mu[X])] + \beta^2 E[X_i + \mu[X]]^2 \\
&= \sigma^2[Y] - 2\beta\sigma[X,Y] + \beta^2\sigma^2[X] \text{-----} 6.5.3
\end{aligned}$$

Substituting Equation 6.3 in Equation 6.5.3 and rewriting the covariance term:

$$E[\varepsilon^2] = \sigma^2[Y] - 2\rho(\sigma[Y]/\sigma[X])\rho\sigma[Y]\sigma[X] + \rho^2(\sigma^2[Y]/\sigma^2[X])\sigma^2[X]$$

$$\begin{aligned}
&= \sigma^2[Y] - 2\rho\sigma[Y]\rho\sigma[Y] + \rho^2\sigma^2[Y] \\
&= \sigma^2[Y] - 2\rho^2\sigma^2[Y] + \rho^2\sigma^2[Y] \\
&= \sigma^2[Y] - \rho^2\sigma^2[Y] \\
&= \sigma^2[Y](1-\rho^2[X,Y]) \text{-----} 6.5.4
\end{aligned}$$

Rewrite Equation 6.5.4 as:

$$\sigma^2[Y|X] = \sigma^2[Y](1-\rho^2[X,Y]) \text{-----} 6.5$$

Equation 6.5 only applies if the entire sub-population of Y values would be available for each fixed X. An unbiased sample estimate of the conditional error variance is obtained through  $s^2[Y|X] = [n/(n-2)][\sigma^2[Y|X]]$ . The square root of  $\sigma^2[Y|X]$  [and  $\sqrt{s^2[Y|X]}$ ] is referred to as the standard error of estimate [Benson, Levine & Goldstein, 1983; Cohen & Cohen, 1983; Tabachnick & Fidell, 1989; Younger, 1979;].

Even though the regression model shown as Equation 6.4 can not account for all the variance in Y in terms of the linear relationship with X [unless  $\rho[X,Y] = 1$  and the linear model provides a perfect fit], it still provides the best possible estimate of Y given X in the sense that the sum of the squared deviations around the regression line is a minimum.

Assuming homoscedasticity, Equation 6.4 permits the estimation of the probability of success [or failure] conditional on X [i.e.  $P[Y \geq Y_c | X = X_i]$ ] as shown below. Assuming the conditional criterion  $[\mu[Y|X], \sigma^2[Y|X]]$  distribution to be normal,  $Y_c$  can be transformed to a Z-score in the conditional criterion distribution through Equation 6.6.

$$Z = (Y_c - E[Y|X=X_i])/\sigma[Y|X] \text{-----} 6.6$$

The probability  $P[Z_Y \geq Z_{Y_c} | X = X_i]$  can subsequently be determined through the integration of the standard normal density function, or alternatively, through the appropriate statistical tables or computer procedures. Figure 6.1 presents a graphical description of the standard error of estimate and its role in the conditional probability of success.

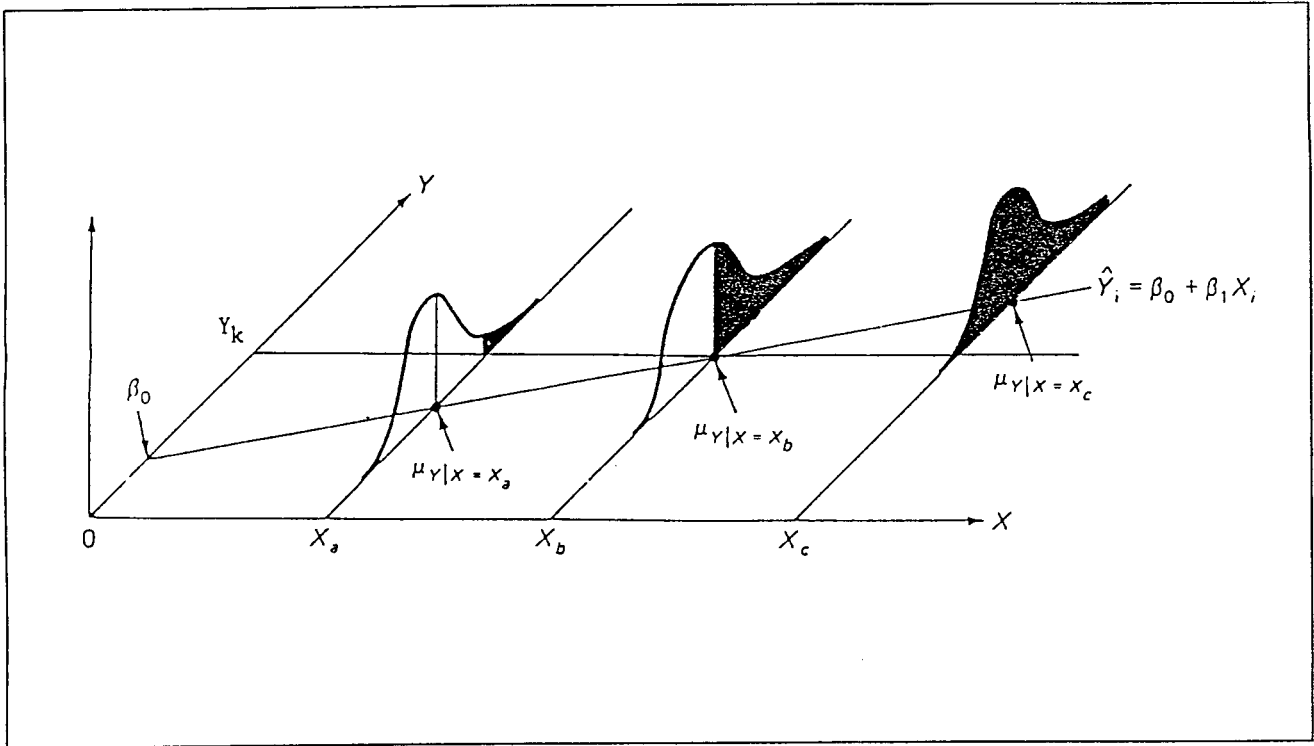


Figure 6.1: The estimation of the probability of success conditional on  $X$  [i.e.  $P[\hat{Y} \geq Y_k | X = X_i]$ ]

## 6.2 EFFECT OF STATISTICAL CORRECTIONS FOR RESTRICTION OF RANGE AND/OR RANDOM MEASUREMENT ERROR IN THE CRITERION ON THE DECISION FUNCTION

The purpose of selection validation research is to formulate and justify an effective and equitable decision rule based on estimates of applicants actual job performance [and not an indicator of it distorted by random measurement error] across the full spectrum of the criterion distribution. Stated differently, the purpose of selection validation research is to formulate and justify an effective and equitable selection decision rule on a sample of applicants from the applicant population that successfully generalises to the actual area of application. If, however, the context in which the decision rule is formulated and justified differs in one or more respects from the context in which the decision rule is meant to be applied, the transferability of both the decision rule and its justification in terms of effectiveness and equity is jeopardised.

The rationale for the previously discussed corrections to the correlation [i.e. validity] coefficient for random measurement error in the criterion and/or restriction of range stem from the necessity of aligning the contexts of development and application. But, as was argued earlier, restricting the corrections to the validity coefficient only would still leave the instructions contained in the strategy

matrix unaffected. The decision strategy actually applied for decision making would thus still be the one derived from a simulated application, which, however, is not fully representative of the actual application. This naturally should lead to the question whether the aforementioned factors that produce bias in the validity coefficient obtained from a simulated application in the development environment also systematically affects the parameters of the decision function [specifically,  $\alpha$ ,  $\beta$ ,  $E[Y | X_i]$  &  $P[Y \geq Y_c | X = X_i]$ ]. If in fact one or more of these situational characteristics do introduce bias in one or more of these parameters, the logical further question arises how to correct for the biasing effect of random measurement error in the criterion and/or restriction of range comparable to the various previous corrections to the correlation [i.e. validity] coefficient. Finally, should such corrections prove to be practical, the following additional questions are raised:

- would such corrections change decisions on employment applicants?; and if so
- what would the value of the consequences of the change in decisions be?; and
- how would the cost of the corrections compare to the value of the changed decisions affected by the corrections?.

The fundamental a priori position underlying these latter questions is that such corrections would serve very little, if any, practical purpose if they do not change the actual decisions on applicants for employment or do so at a cost exceeding the value of the change in decisions.

## 6.2.1 Effect Of Statistical Corrections For Random Measurement Error On The Decision Function

### 6.2.1.1 Effect Of Statistical Corrections For Random Measurement Error In The Criterion On The Decision Function

The effect of removal of measurement error from the criterion only on the parameters of the simple linear model is examined in Equation 6.7.

The expected true criterion score, conditional on the observed predictor score can be written as:

$$E[T_Y | X_i] = \alpha_1 + \beta_1 X_i \text{-----} 6.7.1$$

Where;

$$\alpha_1 = E[T_Y] - \beta_1 E[X]$$

$$= \mu[Y] - \beta_1 \mu[X] \text{-----} 6.7.2$$

and:

$$\begin{aligned} \beta_1 &= \rho[X, T_Y](\sigma[T_Y]/\sigma[X]) \\ &= (\rho[X, Y]/\sqrt{\rho_{tt}Y})\{\sqrt{\rho_{tt}Y}(\sigma[Y]/\sigma[X])\} \\ &= \rho[X, Y](\sigma[Y]/\sigma[X]) \text{-----} 6.7.3 \end{aligned}$$

Substituting Equations 6.7.2 and 6.7.3 in Equation 6.7.1:

$$\begin{aligned} E[T_Y | X_i] &= (\mu[Y] - \beta_1 \mu[X]) + \beta_1 X_i \\ &= \mu[Y] + \beta_1 (X_i - \mu[X]) \\ &= \mu[Y] + \rho[X, Y](\sigma[Y]/\sigma[X])(X_i - \mu[X]) \\ &= E[Y | X_i] \text{-----} 6.7 \end{aligned}$$

Equation 6.7 indicates that the decision function remains unaffected by the partial correction for attenuation. Thus neither the Y-axis intercept nor the X-axis slope are affected by the removal of random measurement error from the criterion. The position of applicants in the strategy matrix, consequently, will also remain unaffected since  $E[T_Y | X] = E[Y | X]$ . The decisions on applicants will therefore not change due to the partial correction for attenuation.

Equation 6.8 points to the same conclusion by indicating that  $E[T_Y | X]$  and  $E[Y | X]$ , standardised on the original, fallible criterion distribution, coincide.

Let  $Z_1$  represent the standardised  $E[Y | X]$

Let  $Z_2$  represent  $E[T_Y | X]$  standardised on the fallible criterion distribution

Expressing the deviation of the expected observed criterion score conditional on the observed predictor score from the criterion mean in terms of the standard deviation of the observed criterion distribution:

$$Z_1 = (E[Y | X] - \mu[Y])/\sigma[Y] \text{-----} 6.8.1$$

Substituting Equation 6.1 in Equation 6.8.1:

$$Z_1 = (\alpha + \beta[Y | X]X - \mu[Y])/\sigma[Y] \text{-----} 6.8.2$$



Substituting equation 6.2 in Equation 6.8.2:

$$\begin{aligned} Z_1 &= (\mu[Y] - \beta[Y|X]\mu[X] + \beta[Y|X]X - \mu[Y])/\sigma[Y] \\ &= \beta[Y|X](X - \mu[X])/\sigma[Y] \end{aligned} \quad \text{6.8.3}$$

Substituting Equation 6.3 in Equation 6.8.3:

$$\begin{aligned} Z_1 &= (\rho[X,Y]\sigma[Y](X - \mu[X]))/(\sigma[X]\sigma[Y]) \\ &= (\rho[X,Y](X - \mu[X]))/\sigma[X] \end{aligned} \quad \text{6.8.4}$$

Similarly, expressing the deviation of the expected true criterion score conditional on the observed predictor score from the criterion mean in terms of the standard deviation of the observed criterion distribution:

$$\begin{aligned} Z_2 &= (E[T_Y|X] - \mu[Y])/\sigma[Y] \\ &= (\alpha[T_Y|X] - \beta[T_Y|X] - \mu[Y])/\sigma[Y] \end{aligned} \quad \text{6.8.5}$$

Substituting Equation 6.7 in Equation 6.8.5:

$$\begin{aligned} Z_2 &= \{\mu[Y] + \rho[X,Y](\sigma[Y]/\sigma[X])(X_i - \mu[X]) - \mu[Y]\}/\sigma[Y] \\ &= (\rho[X,Y](X - \mu[X]))/\sigma[X] \end{aligned} \quad \text{6.8.6}$$

Therefore:

$$\begin{aligned} Z_1/Z_2 &= \{(\rho[X,Y](X - \mu[X]))/\sigma[X]\} \{\sigma[X]/(\rho[X,Y](X - \mu[X]))\} \\ &= 1 \end{aligned} \quad \text{6.8}$$

Therefore:

$$Z_1 = Z_2$$

The critical criterion cutoff value,  $Y_c$ , is defined on a scale unaffected by [systematic and/or random] measurement error.  $Y_c$  thus reflects the minimum actual contribution considered acceptable. Equation 6.9 indicates that the removal of random measurement error from the criterion consequently has the effect of increasing the absolute value of  $Y_c$  expressed as a Z-score standardised on the infallible criterion distribution.  $Y_c$  thus effectively shifts away from the mean as a function of  $\rho_{ttY}$  thereby reducing the proportion of the criterion distribution falling above  $Y_c$  [i.e. the base rate] if  $Y_c > \mu[Y]$  or increasing the base rate if  $Y_c < \mu[Y]$ .

Let  $Z_1$  represent the standardised criterion cutoff in the original criterion distribution

let  $Z_2$  represent the standardised criterion cutoff in the disattenuated criterion distribution

The criterion cutoff standardised on the original criterion distribution can be expressed as:

$$Z_1 = (Y_c - \mu[Y]) / \sigma[Y] \text{-----} 6.9.1$$

The criterion cutoff standardised on the true score criterion distribution can be expressed as:

$$Z_2 = (Y_c - \mu[Y]) / (\sigma[Y] \sqrt{\rho_{tt}Y}) \text{-----} 6.9.2$$

Therefore:

$$\begin{aligned} Z_1/Z_2 &= ((Y_c - \mu[Y]) / \sigma[Y]) (\sigma[Y] \sqrt{\rho_{tt}Y}) / (Y_c - \mu[Y]) \\ &= \sqrt{\rho_{tt}Y} \\ &\leq 1 \text{-----} 6.9 \end{aligned}$$

Therefore:

$$|Z_1| \leq |Z_2|$$

And;

$$Z_2 = \sqrt{\rho_{tt}Y} [Z_1]$$

Therefore:

$$BR_1 \geq BR_2$$

Although the expected criterion performance conditional on  $X$  is not affected by the partial correction for attenuation, the relative position of the expected performance in the criterion distribution does change as shown by Equation 6.10. Comparing Equations 6.9 and 6.10, however, indicates that the change in the position of  $Y_c$  relative to the change in the position of the expected criterion performance is of the same magnitude and in the same direction. The corrections, therefore, do not affect the position of the expected performance relative to  $Y_c$ . The decisions on applicants will therefore not change due to the partial correction for attenuation.

Let  $Z_1$  represent the expected observed score criterion performance conditional on  $X_i$   $E[Y|X_i]$  standardised on the observed score criterion distribution.

Let  $Z_2$  represent the expected true score criterion performance conditional on  $X_i$   $E[T_Y|X_i]$  standardised on the true score criterion distribution.

According to Equation 6.8.4, the expected observed score criterion performance conditional on  $X_i$  standardised on the observed score criterion distribution can be expressed as:

$$\begin{aligned} Z_1 &= (E[Y|X_i] - \mu[Y])/\sigma[Y] \\ &= (\rho[X,Y]\sigma[Y](X - \mu[X]))/(\sigma[X]\sigma[Y]) \\ &= (\rho[X,Y](X - \mu[X]))/\sigma[X] \end{aligned} \quad \text{6.10.1}$$

The expected true score criterion performance conditional on  $X_i$ , standardised on the true score criterion distribution can be expressed as:

$$\begin{aligned} Z_2 &= (E[T_Y|X] - \mu[Y])/\sigma[T_Y] \\ &= (E[T_Y|X] - \mu[Y])/(\sigma[Y]\sqrt{\rho_{tt}Y}) \end{aligned} \quad \text{6.10.2}$$

Substituting Equation 6.7 in Equation 6.10.2:

$$\begin{aligned} Z_2 &= \{\mu[Y] + \rho[X,Y](\sigma[Y]/\sigma[X])(X_i - \mu[X]) - \mu[Y]\}/(\sigma[Y]\rho_{tt}Y) \\ &= \{\rho[X,Y](X_i - \mu[X])\}/(\sigma[X]\sqrt{\rho_{tt}Y}) \end{aligned} \quad \text{6.10.3}$$

Therefore:

$$\begin{aligned} Z_1/Z_2 &= \{(\rho[X,Y](X_i - \mu[X]))/\sigma[X]\} \{(\sigma[X]\sqrt{\rho_{tt}Y})/(\rho[X,Y](X_i - \mu[X]))\} \\ &= \sqrt{\rho_{tt}Y} \\ &\leq 1 \end{aligned} \quad \text{6.10.4}$$

Therefore:

$$|Z_1| \leq |Z_2|$$

The preceding argument, however, applies only if the information way of the strategy matrix is expressed in terms of expected criterion performance. Should the information way of the strategy matrix be expressed in terms of the conditional probability of success, attention turns to the standard error of estimate as the critical factor [assuming  $Y_c$  to remain unaffected]. Equation 6.11 reflects the

effect of statistically removing random measurement error from the criterion on the standard error of estimate.

According to Equation 6.5:

$$\sigma^2[Y | X] = \sigma^2[Y](1 - \rho^2[X, Y]) \text{-----} 6.11.1$$

Therefore:

$$\begin{aligned} \sigma^2[T_Y | X] &= \sigma^2[T_Y](1 - \rho^2[X, T_Y]) \\ &= \rho_{ttY} \sigma^2[Y] \{1 - (\rho^2[X, Y] / \rho_{ttY})\} \\ &= \rho_{ttY} \sigma^2[Y] - \rho_{ttY} \sigma^2[Y] (\rho^2[X, Y] / \rho_{ttY}) \\ &= \rho_{ttY} \sigma^2[Y] - \sigma^2[Y] \rho^2[X, Y] \\ &= \sigma^2[Y] (\rho_{ttY} - \rho^2[X, Y]) \\ &\leq \sigma^2[Y | X] \text{-----} 6.11 \end{aligned}$$

Therefore:

$$\sigma[T_Y | X] \leq \sigma[Y | X]$$

The effect of statistically removing random measurement error from the criterion is to reduce the standard error of estimate. The critical question is how this affects the conditional probability of success. Assuming a constant  $Y_c$ , only the reaction of the standard of estimate to the removal of random measurement error from the criterion needs to be taken into consideration since the parameters of the decision function [i.e. the expected criterion performance] remain invariant. The decrease in the standard error brought about by the removal of measurement error from the criterion, will have the effect of translating the critical criterion cutoff to a more extreme Z-score as shown by Equation 6.12.

Let  $Z_1$  represent the standardised  $Y_c$  in the fallible conditional criterion distribution

Let  $Z_2$  represent the standardised  $Y_c$  in the infallible conditional criterion distribution

The critical criterion cutoff  $Y_c$  standardised on the fallible conditional criterion distribution can be expressed as:

$$Z_1 = (Y_c - E[Y | X_i]) / \sigma[Y | X] \text{-----} 6.12.1$$

Substituting Equation 6.5 in Equation 6.12.1:

$$Z_1 = (Y_c - E[Y | X_i]) / \{\sigma[Y] \sqrt{(1 - \rho^2[X, Y])}\} \text{-----} 6.12.2$$

The critical criterion cutoff  $Y_c$  standardised on the infallible conditional criterion distribution can be expressed as:

$$Z_2 = (Y_c - E[T_Y | X_i]) / \sigma[T_Y | X] \text{-----} 6.12.3$$

Substituting Equations 6.7 and 6.11 in Equation 6.12.3:

$$Z_2 = (Y_c - E[Y | X_i]) / (\sigma[Y] \sqrt{\rho_{ttY} \rho^2[X, Y]}) \text{-----} 6.12.4$$

Therefore:

$$\begin{aligned} Z_1/Z_2 &= \{(Y_c - E[Y | X_i]) / \sigma[Y] \sqrt{(1 - \rho^2[X, Y])}\} \{ \sigma[Y] \sqrt{\rho_{ttY} \rho^2[X, Y]} / (Y_c - E[Y | X_i]) \} \\ &= (\sqrt{\rho_{ttY} \rho^2[X, Y]} / \sqrt{(1 - \rho^2[X, Y])}) \\ &\leq 1 \text{-----} 6.12 \end{aligned}$$

Therefore:

$$|Z_1| \leq |Z_2|$$

The precise effect on  $P[T_Y \geq Y_c | X = X_i]$ , moreover, will depend on the position of  $X_i$  relative to  $X_c = (Y_c - \alpha) / \beta$  [i.e.  $P[Y \geq Y_c | X = X_c] = 0.50$ ]. For all  $X_i < X_c$ ,  $P[T_Y \geq Y_c | X = X_i]$  will decrease relative to  $P[Y \geq Y_c | X = X_i]$  and for all  $X_i > X_c$ ,  $P[T_Y \geq Y_c | X = X_i]$  will increase relative to  $P[Y \geq Y_c | X = X_i]$ . For  $X_i = X_c$ ,  $P[Y \geq Y_c | X = X_i]$  will remain unaffected.

Partially correcting the standard error of estimate for the attenuating effect of criterion unreliability will, however, very unlikely change the selection decisions on applicants when selection is restricted by quotas. Although the conditional probabilities are altered by the correction, the rank-order of the applicants in terms of their chances of success remain exactly the same. Consequently, the same top  $N_v$  would still be selected. Partially correcting the standard error of estimate for the attenuating effect of criterion unreliability will only affect restricted selection if an insufficient number of applicants initially [i.e. prior to corrections] meet the entry requirement  $\alpha_s$ .

In the case of unrestricted selection, however, partially correcting the standard error for criterion unreliability could affect the selection decision on applicants by increasing the number of applicants

qualifying for selection [assuming  $\alpha_s > 0.50$ ]. Correcting the standard error for criterion unreliability has the effect of pushing any conditional probability greater than 0.50 but less than  $\alpha_s$  [assuming  $\alpha_s > 0.50$ ] towards, and possibly past,  $\alpha_s$ , thus increasing the selection ratio. Correcting the standard error for criterion unreliability has the effect of pushing the critical predictor cutoff corresponding to the critical probability  $\alpha_s$  towards the point  $X_c$  [as defined above] in the predictor distribution. The selection ratio is consequently increased for any  $\alpha_s > 0.50$  and decreased for any  $\alpha_s < 0.50$  [although the chance of such a liberal entry requirement seems unlikely].

### 6.2.1.2 Effect Of Statistical Corrections For Random Measurement Error In Both The Criterion And The Predictor On The Decision Function

The effect of removal of measurement error from both the criterion and the predictor on the parameters of the simple linear model is examined in Equation 6.13.

The linear regression of the true criterion scores on the true predictor scores can be expressed as:

$$E[T_Y | T_X] = \alpha[T_Y | T_X] + \beta[T_Y | T_X]T_X \text{-----} 6.13.1$$

The intercept parameter in Equation 6.13.1 can be expressed as:

$$\begin{aligned} \alpha[T_Y | T_X] &= E[T_Y] - \beta[T_Y | T_X]E[T_X] \\ &= \mu[Y] - \beta[T_Y | T_X]\mu[X] \text{-----} 6.13.2 \end{aligned}$$

The slope parameter in Equation 6.13.1 can be written as:

$$\beta[T_Y | T_X] = \rho[T_X, T_Y](\sigma[T_Y]/\sigma[T_X]) \text{-----} 6.13.3$$

Substituting Equations 3.1 and 2.16 in Equation 6.13.3:

$$\begin{aligned} \beta[T_Y | T_X] &= \{\rho[X, Y]/(\sqrt{\rho_{ttY}}\sqrt{\rho_{ttX}})\}\{\sqrt{\rho_{ttY}}\sigma[Y]/(\sqrt{\rho_{ttX}}\sigma[X])\} \\ &= (\rho[X, Y](\sigma[Y]/\sigma[X])(1/\sqrt{\rho_{ttX}}) \\ &= (1/\rho_{ttX})\beta[Y | X] \text{-----} 6.13.4 \end{aligned}$$

Therefore:

$$E[T_Y | T_{Xi}] = \mu[Y] + (\rho[X, Y]/\rho_{ttX})(\sigma[Y]/\sigma[X])(T_{Xi} - \mu[X])$$



$$\begin{aligned} &\neq E[Y | X_i] \\ &\neq E[T_Y | X_i] \end{aligned} \text{-----} 6.13$$

Equation 6.13 indicates that the decision function derived from infallible [i.e. perfectly reliable] predictor and criterion data differs from the decision function derived from fallible criterion and predictor data. The effect of the removal of random measurement error from both the dependent and independent variables is to increase the slope parameter  $\beta$  and to decrease the intercept parameter  $\alpha$  as a function of  $\rho_{ttX}$ . This implies that the regression equation derived from fallible data and the regression equation derived from infallible data will intersect at some point  $X_s$ . The point of intersection  $X_s$  coincides with the mean of the predictor distributions [i.e.  $\mu[X] = \mu[T_X]$ ] since  $E[T_Y | T_X = \mu[T_X]] = \mu[Y] = E[Y | X = \mu[X]]$ .

The problem with the fully disattenuated correlation coefficient, and thus also with Equation 6.13, in the context of validation research lies in the fact that a perfectly reliable predictor will never be available.  $T_X$  can therefore never be obtained directly. The actual effect of substituting the regression equation derived from fallible data [Equation 6.4] with an equation derived from infallible data [Equation 6.13] can thus never be assessed since the reaction of individual X-values to the removal of their random measurement error component can never be ascertained.

Although  $T_X$  can never be directly calculated, it can nonetheless be estimated through Equation 6.14, albeit with some error.

The linear regression of the true predictor scores on the observed predictor scores can be expressed as:

$$E[T_X | X] = \alpha[T_X | X] + \beta[T_X | X]X_i \text{-----} 6.14.1$$

The intercept parameter in equation 6.14.1 can be expressed as:

$$\begin{aligned} \alpha[T_X | X] &= \mu[T_X] - \beta[T_X | X]\mu[X] \\ &= \mu[X] - \beta[T_X | X]\mu[X] \end{aligned} \text{-----} 6.14.2$$

Since  $\mu[T_X] = \mu[X]$

The slope parameter in Equation 6.14.1 can be expressed as:

$$\begin{aligned} \beta[T_X | X] &= \rho[T_X, X](\sigma[T_X]/\sigma[X]) \\ &= \sqrt{\rho_{ttX}}\sqrt{\rho_{ttX}} \\ &= \rho_{ttX} \end{aligned} \text{-----} 6.14.4$$

Substituting Equations 6.14.2 and 6.14.4 in Equation 6.14.1:

$$\begin{aligned} E[T_X | X] &= \mu[X] - \beta[T_X | X]\mu[X] + \beta[T_X | X]X_i \\ &= \mu[X](1 - \beta[T_X | X]) + \beta[T_X | X]X_i \\ &= \mu[X](1 - \rho_{ttX}) + \rho_{ttX}X_i \end{aligned} \quad \text{----- 6.14}$$

Substituting Equation 6.14 into Equation 6.13 results in Equation 6.15 returning the expected actual criterion performance conditional on the expected true predictor score, conditional on the observed predictor score. Obtaining an expression for  $E[T_Y | E[T_X | X]]$ , however, requires more than simply exchanging  $T_X$  in Equation 6.13 with Equation 6.14. To do so would result in an inaccurate argument although it would still have led to the same eventual conclusion that  $E[T_Y | E[T_X | X]] = E[Y | X_i] = E[T_Y | X_i]$ . To obtain an appropriate expression for  $E[T_Y | E[T_X | X]]$  requires, in addition, that  $\rho[T_Y, T_X]$  and  $\sigma[T_X]$  be replaced by  $\rho[T_Y, E[T_X | X]]$  and  $\sigma[E[T_X | X]]$  respectively.

The regression of the true criterion scores on the expected true predictor scores conditional on the observed predictor scores, can be expressed as:

$$E[T_Y | E[T_X | X]] = \alpha[T_Y | E[T_X | X]] + \beta[T_Y | E[T_X | X]]E[T_X | X] \quad \text{----- 6.15.1}$$

The intercept parameter in Equation 6.15.1 can be expressed as:

$$\begin{aligned} \alpha[T_Y | E[T_X | X]] &= \mu[T_Y] - \beta[T_Y | E[T_X | X]]\mu[E[T_X | X]] \\ &= \mu[Y] - \beta[T_Y | E[T_X | X]]\mu[X] \end{aligned} \quad \text{----- 6.15.2}$$

The slope parameter in Equation 6.15.1 can be expressed as:

$$\begin{aligned} \beta[T_Y | E[T_X | X]] &= \rho[T_Y, E[T_X | X]](\sigma[T_Y]/\sigma[E[T_X | X]]) \\ &= \rho[T_Y, X]\{(\sigma[Y]\sqrt{\rho_{ttY}})/(\sigma[X]\rho_{ttX})\} \\ &= (\rho[X, Y]/\sqrt{\rho_{ttY}})(\{\sigma[Y]\sqrt{\rho_{ttY}}\}/(\sigma[X]\rho_{ttX})) \\ &= (\rho[X, Y]\sigma[Y]) / (\sigma[X]\rho_{ttX}) \\ &= (1/\rho_{ttX})\beta[Y | X] \end{aligned} \quad \text{----- 6.15.3}$$

Since,  $\rho[T_Y, E[T_X | X]] = \rho[T_Y, X]$  since  $T_X$  is a linear function of  $X$

and

Since,  $\sigma[E[T_X | X]] = \sqrt{(\rho^2[T_X, X]\sigma^2[T_X])} = \sqrt{(\rho_{ttX}\rho_{ttX}\sigma^2[X])} = \rho_{ttX}\sigma[X]$

Substituting Equation 6.15.2 in Equation 6.15.1:

$$\begin{aligned} E[T_Y | E[T_X | X]] &= \alpha[T_Y | E[T_X | X]] + \beta[T_Y | E[T_X | X]]E[T_X | X] \\ &= \mu[Y] - \beta[T_Y | E[T_X | X]]\mu[X] + \beta[T_Y | E[T_X | X]]E[T_X | X] \\ &= \mu[Y] + \beta[T_Y | E[T_X | X]](E[T_X | X] - \mu[X]) \text{-----} 6.15.4 \end{aligned}$$

Substituting Equation 6.15.3 in Equation 6.15.4:

$$\begin{aligned} E[T_Y | E[T_X | X]] &= \mu[Y] + \beta[T_Y | E[T_X | X]](E[T_X | X] - \mu[X]) \\ &= \mu[Y] + (1/\rho_{tX})\beta[Y | X](E[T_X | X] - \mu[X]) \\ &= \mu[Y] + (1/\rho_{tX})\{\rho[X, Y](\sigma[Y]/\sigma[X])\}(E[T_X | X] - \mu[X]) \text{-----} 6.15 \end{aligned}$$

Equation 6.15, similar to Equation 6.13, differs from the fallible regression equation and the partially attenuated regression equation [Equation 6.4] in terms of both intercept and slope. The regression of  $T_Y$  on  $E[T_X | X]$  has a steeper slope and a lower intercept than the regression of  $Y$  on  $X$ . The increase in slope is a function of  $1/\rho_{tX}$ ; thus the less reliable the predictor measurements the greater the increase in the slope and the corresponding decrease in intercept. The regression of  $T_Y$  on  $E[T_Y | X]$ , furthermore, corresponds to the regression of  $T_Y$  on  $T_X$  in terms of intercept and slope.

Since it is possible to obtain the regression of  $T_Y$  on  $T_X$  from the regression of  $T_Y$  on  $E[T_X | X]$ , the temptation exists to conclude that the possibility presents itself to obtain a closer estimation of  $E[T_Y | T_X]$  than is possible through  $E[T_Y | X]$ . One should, however, be careful in concluding that the change in the decision function implies a concomitant change in the expected criterion performance associated with the various applicants. In assessing the effect of fully disattenuating the parameters of the decision function, the regression of  $E[T_X | X]$  and  $E[T_Y | Y]$  towards their respective means must be kept in mind. The change in the predictor information entering the regression equation and the change in the regression parameters must consequently be considered simultaneously to assess the change in the expected criterion performance attributable to Equation 6.15.

Let  $Z_1$  represent the expected criterion performance obtained from the fallible regression equation [Equation 6.4], standardised on the fallible criterion distribution.

Let  $Z_2$  represent the expected criterion performance obtained from Equation 6.15 standardised on the fallible criterion distribution.

The expected observed criterion performance, conditional on  $X_i$ , standardised on the fallible criterion distribution can be expressed as:

$$Z_1 = (E[Y | X_i] - \mu[Y]) / \sigma[Y] \text{-----} 6.16.1$$

Substitute Equations 6.1, 6.2 and 6.3 in Equation 6.16.1:

$$\begin{aligned} Z_1 &= (\alpha + \beta[Y | X]X_i - \mu[Y]) / \sigma[Y] \\ &= (\mu[Y] - \beta[Y | X]\mu[X] + \beta[Y | X]X_i - \mu[Y]) / \sigma[Y] \\ &= \beta[Y | X](X_i - \mu[X]) / \sigma[Y] \\ &= (\rho[X, Y]\sigma[Y]) / \{\sigma[X](X_i - \mu[X]) / \sigma[Y]\} \\ &= (\rho[X, Y] / \sigma[X])(X_i - \mu[X]) \text{-----} 6.16.2 \end{aligned}$$

The expected true score criterion performance obtained from Equation 6.15 standardised on the fallible criterion distribution can be expressed as:

$$Z_2 = (E[T_Y | E[T_X | X]] - \mu[Y]) / \sigma[Y] \text{-----} 6.16.3$$

Substituting Equations 6.15.1 and 6.15.2 in Equation 6.16.3

$$\begin{aligned} Z_2 &= (\alpha + \beta[T_Y | E[T_X | X]]E[T_X | X_i] - \mu[Y]) / \sigma[Y] \\ &= (\mu[Y] - \beta[T_Y | E[T_X | X]]\mu[X] + \beta[T_Y | E[T_X | X]]E[T_X | X_i] - \mu[Y]) / \sigma[Y] \\ &= \beta[T_Y | E[T_X | X]](E[T_X | X_i] - \mu[X]) / \sigma[Y] \text{-----} 6.16.4 \end{aligned}$$

Substituting Equation 6.15.3 in Equation 6.16.4:

$$\begin{aligned} Z_2 &= \{(\rho[X, Y]\sigma[Y]) / (\sigma[X]\rho_{ttX})\}(E[T_X | X_i] - \mu[X]) / \sigma[Y] \\ &= \{(\rho[X, Y])(E[T_X | X_i] - \mu[X])\} / (\sigma[X]\rho_{ttX}) \text{-----} 6.16.5 \end{aligned}$$

Substituting Equation 6.14 in Equation 6.16.5:

$$\begin{aligned} Z_2 &= \{(\rho[X, Y])((\mu[X](1 - \rho_{ttX}) + \rho_{ttX}X_i) - \mu[X])\} / (\sigma[X]\rho_{ttX}) \\ &= \{(\rho[X, Y])(\mu[X] - \mu[X]\rho_{ttX} + \rho_{ttX}X_i - \mu[X])\} / (\sigma[X]\rho_{ttX}) \\ &= \{(\rho[X, Y])(\mu[X]\rho_{ttX} + \rho_{ttX}X_i)\} / (\sigma[X]\rho_{ttX}) \\ &= \{(\rho[X, Y])(\rho_{ttX}(X_i - \mu[X]))\} / (\sigma[X]\rho_{ttX}) \\ &= \{(\rho[X, Y])(X_i - \mu[X])\} / (\sigma[X]) \text{-----} 6.16.6 \end{aligned}$$

Therefore:

$$\begin{aligned} Z_1 / Z_2 &= (\rho[X, Y] / \sigma[X])(X_i - \mu[X])(\sigma[X]) / \{\rho[X, Y](X_i - \mu[X])\} \\ &= 1 \text{-----} 6.16 \end{aligned}$$

Therefore:

$$|Z_1| = |Z_2|$$

Equation 6.16 indicates that the corrections have no effect on the expected criterion performance. The expected  $T_Y$  performance derived through Equation 6.15 equals the expected criterion performance derived through the uncorrected regression of  $Y$  on  $X$  or partially corrected regression of  $T_Y$  on  $X$ .

The removal of random measurement error from both the dependent and independent variables will consequently have no affect on the applicants selected irrespective of whether selection is restricted by quotas or not and strict top down selection applies..

The issue can, however, also be approached from the perspective of the predictor. Assume a critical predictor cutoff  $X_C = (Y_C - \alpha[Y | X]) / \beta[Y | X]$ , so that  $E[Y | X = X_C] = Y_C$  and assume a critical expected true score  $X_{CC} = (Y_C - \alpha[Y | E[T_X | X]]) / \beta[T_Y | E[T_X | X]]$ , so that  $E[T | E[T_X | X] = X_C] = Y_C$ .

Equation 6.17 indicates that the ratio of the standardised  $X_C$  [ $Z_1$ ] to  $X_{CC}$  standardised on the original predictor distribution [ $Z_2$ ] always will equal or exceed unity, which implies that  $X_{CC}$  will either coincide with  $X_C$  or occupy a less extreme position than  $X_C$  in the original predictor distribution.

The critical predictor cutoff derived from the equation describing the regression of the fallible criterion scores on the fallible predictor score [Equation 6.4], can be expressed as:

$$\begin{aligned} X_C &= (Y_C - \alpha[Y | X]) / \beta[Y | X] \\ &= (Y_C - (\mu[Y] - \beta[Y | X]\mu[X])) / \beta[Y | X] \\ &= Y_C / \beta[Y | X] - \mu[Y] / \beta[Y | X] + \mu[X] \\ &= (1 / \beta[Y | X])(Y_C - \mu[Y]) + \mu[X] \\ &= (\sigma[X] / (\rho[X, Y]\sigma[Y]))(Y_C - \mu[Y]) + \mu[X] \end{aligned} \quad \text{----- 6.17.1}$$

Similarly from Equation 6.15 follows that the critical predictor cutoff derived from the equation describing the regression of the true criterion scores on the estimated true predictor scores, can be expressed as:

$$\begin{aligned} X_{CC} &= 1 / \beta[T_Y | E[T_X | X]](Y_C - \mu[Y]) + \mu[X] \\ &= (\sigma[X] \rho_{tTX} / (\rho[X, Y]\sigma[Y]))(Y_C - \mu[Y]) + \mu[X] \end{aligned} \quad \text{----- 6.17.2}$$

Let  $Z_1$  represent the standardised  $X_C$

Let  $Z_2$  represent  $X_{cc}$  standardised in terms of the fallible predictor distribution

Consequently:

$$\begin{aligned} Z_1/Z_2 &= ((X_c - \mu[X])/\sigma[X]) / \{(X_{cc} - \mu[X])/\sigma[X]\} \\ &= ((X_c - \mu[X])/\sigma[X])(\sigma[X]/(X_{cc} - \mu[X])) \\ &= ((X_c - \mu[X])/(X_{cc} - \mu[X])) \end{aligned} \quad 6.17.3$$

Substituting Equations 6.17.1 and 6.17.2 in Equation 6.17.3:

$$\begin{aligned} Z_1/Z_2 &= \{(\sigma[X]/(\rho[X,Y]\sigma[Y]))(Y_c - \mu[Y]) + \\ &\quad \mu[X] - \mu[X]\} / \{(\sigma[X]\rho_{ttX}/(\rho[X,Y]\sigma[Y]))(Y_c - \mu[Y]) + \mu[X] - \mu[X]\} \\ &= \{\sigma[X]/(\rho[X,Y]\sigma[Y])\}(Y_c - \mu[Y])(\rho[X,Y]\sigma[Y])(1/\sigma[X]\rho_{ttX})(1/(Y_c - \mu[Y])) \\ &= 1/\rho_{ttX} \\ &\geq 1 \end{aligned} \quad 6.17$$

Therefore:

$$|Z_1| \geq |Z_2|$$

and:

$$Z_2 = Z_1\rho_{ttX}$$

How this shift in the critical predictor affects the selection ratio will, however, depend on the direction and magnitude of the shift in the predictor information produced by Equation 6.15

Equation 6.18 [see also Equation 6.14] indicates the effect of the removal of random error variance from the predictor and the estimation of the true predictor score on the relative position of applicants in the original predictor distribution. Since  $Z_1/Z_2 \geq 1$  it implies that the estimated true score returned by Equation 6.14 falls numerically closer to the mean of the predictor distribution than the unattenuated, observed score.

Let  $Z_1$  represent the standardised observed predictor score.

Let  $Z_2$  represent the estimated true predictor score standardised on the original predictor distribution.

An observed predictor score, standardised on the observed predictor distribution can be expressed as:



$$Z_1 = (X_i - \mu[X]) / \sigma[X] \text{-----} 6.18.1$$

A true predictor score, estimated from the observed predictor score, standardised on the observed predictor distribution can be expressed as:

$$Z_2 = (E[T_X | X] - \mu[X]) / \sigma[X] \text{-----} 6.18.2$$

Substituting Equation 6.14 in Equation 6.18.2:

$$\begin{aligned} Z_2 &= (\mu[X](1 - \rho_{ttX}) + \rho_{ttX}X_i - \mu[X]) / \sigma[X] \\ &= (\mu[X] - \mu[X]\rho_{ttX} + \rho_{ttX}X_i - \mu[X]) / \sigma[X] \\ &= \rho_{ttX}(X_i - \mu[X]) / \sigma[X] \text{-----} 6.18.13 \end{aligned}$$

Therefore:

$$\begin{aligned} Z_1/Z_2 &= (X_i - \mu[X])(1/\sigma[X])(\sigma[X])(1/\rho_{ttX}(X_i - \mu[X])) \\ &= 1/\rho_{ttX} \\ &\geq 1 \text{-----} 6.18 \end{aligned}$$

Therefore:

$$|Z_1| \geq |Z_2|$$

and:

$$Z_2 = Z_1\rho_{ttX}$$

It thus follows that the magnitude of the  $X_i \rightarrow E[T_X | X_i]$  movement towards the predictor mean equals the magnitude of the  $X_c \rightarrow X_{cc}$  movement towards the mean. Consequently the number of applicants satisfying the condition  $E[T_X | X_i] \geq X_{cc}$  will equal the number of applicants satisfying the condition  $X_i \geq X_c$ .

Selection decisions can be based either on  $E[Y | X]$  or  $P[Y > Y_c | X]$ . The latter is dependent on the former as well as on the conditional criterion variance [i.e., the standard error of estimate]. Equation 6.19 reflects the effect of statistically removing random measurement error from both the criterion and the predictor via Equation 6.15 on the standard error of estimate.

According to Equation 6.5:

$$\sigma^2[Y | X] = \sigma^2[Y](1 - \rho^2[X, Y]) \text{-----} 6.19.1$$

Therefore:

$$\begin{aligned} \sigma^2[T_Y | E[T_X | X_i]] &= \sigma^2[T_Y](1 - \rho^2[T_Y, E[T_X | X_i]]) \\ &= \rho_{ttY} \sigma^2[Y] \{1 - (\rho^2[X, Y] / \rho_{ttY})\} \\ &= \rho_{ttY} \sigma^2[Y] - \rho_{ttY} \sigma^2[Y] (\rho^2[X, Y] / \rho_{ttY}) \\ &= \rho_{ttY} \sigma^2[Y] - \sigma^2[Y] (\rho^2[X, Y]) \\ &= \sigma^2[Y] \{\rho_{ttY} - \rho^2[X, Y]\} \\ &= \sigma^2[T_Y | X] \leq \sigma^2[Y | X] \text{-----} 6.19.2 \end{aligned}$$

Thus:

$$\sigma[T_Y | E[T_X | X_i]] = \sigma[T_Y | X] \leq \sigma[Y | X] \text{-----} 6.19$$

The effect of statistically removing random measurement error from the both the criterion and predictor essentially duplicates the effect of partially correcting the standard error of estimate for criterion unreliability only. The results reported earlier thus also apply to Equation 6.19. Such would not have been the case had it not been for the fact that the parameters of the decision function react to the removal of random measurement error from both the criterion and the predictor in a way that leaves the expected criterion performance unaltered. An analysis of the effect of fully disattenuating the standard error of estimate on the conditional probability of success and eventually selection decision-making would otherwise have had to simultaneously consider the reduction in the standard error of estimate and the change in the parameters of the decision function [i.e.  $E[Y | X_i]$  relative to  $Y_c$ ]. Equation 20, assuming a constant critical criterion cut-off defined on the true criterion scale, indicates that the removal of random error variance from both the criterion and the predictor has the effect of shifting the relative position of  $Y_c$  outwards in the conditional criterion distribution. Equation 6.20 assumes that  $Y_c > \mu[Y]$ .

Let  $Z_1$  represent the standardised  $Y_c$  in the uncorrected conditional criterion distribution

Let  $Z_2$  represent the standardised  $Y_c$  in the corrected conditional criterion distribution

Based on Equation 6.5, the standardised  $Y_c$  in the uncorrected conditional criterion distribution can be expressed as:

$$\begin{aligned} Z_1 &= (Y_c - E[Y | X_i]) / \sigma[Y | X] \\ &= (Y_c - E[Y | X_i]) / \sigma[Y] \sqrt{(1 - \rho^2[X, Y])} \text{-----} 6.20.1 \end{aligned}$$

Based on Equation 6.11, the standardised  $Y_c$  in the corrected conditional criterion distribution in turn can be expressed as:

$$Z_2 = (Y_c - E[T_Y | E[T_X | X_i]]) / \{\sigma[Y] \sqrt{(\rho_{tt}Y - \rho^2[X, Y])}\} \text{-----} 6.20.2$$

Consequently:

$$\begin{aligned} Z_1/Z_2 &= \{(Y_c - E[Y | X_i]) / (\sigma[Y] \sqrt{(1 - \rho^2[X, Y])})\} \{\sigma[Y] \sqrt{(\rho_{tt}Y - \rho^2[X, Y])} / (Y_c - E[T_Y | E[T_X | X_i]])\} \\ &= \{(Y_c - E[Y | X_i]) / \sqrt{(1 - \rho^2[X, Y])}\} \{\sqrt{(\rho_{tt}Y - \rho^2[X, Y])} / (Y_c - E[Y | X_i])\} \\ &= \{\sqrt{(\rho_{tt}Y - \rho^2[X, Y])} / \sqrt{(1 - \rho^2[X, Y])}\} \\ &\leq 1 \text{-----} 6.20 \end{aligned}$$

Since  $(\sqrt{\rho_{tt}Y - \rho^2[X, Y]}) \leq \sqrt{(1 - \rho^2[X, Y])}$

And consequently that the standardised  $Y_c$  in the uncorrected conditional criterion distribution occupies a less extreme position in the fallible criterion distribution than it does in the corrected conditional criterion distribution:

$$|Z_1| < |Z_2|$$

The precise effect of statistically removing random measurement error from both the predictor and the criterion on  $P[T_Y \geq Y_c | E[T_X | X]]$  will depend on the position of  $T_{X_i}$  relative to  $X_{cc}$  [or  $X_i$  relative to  $X_c$ ]. The fully disattenuated conditional probability of success will increase relative to  $P[Y \geq Y_c | X]$  for all applicants with expected criterion performances greater than the critical criterion cut-off while it would decrease relative to  $P[T \geq Y_c | X]$  for all applicants with  $E[Y | X] < Y_c$ .  $P[T_Y \geq Y_c | E[T_X | X]]$  will consequently provide a closer approximation of  $P[T_Y \geq Y_c | T_X]$  than the attenuated conditional probability of success.

The fully disattenuated standard error of measurement applicable to Equation 6.13 could likewise be determined but would, like Equation 6.13, have relatively little practical utility since  $E[T_Y | T_X]$  can not be estimated and hence  $P[T_Y \geq Y_c | T_X]$  will remain an unattainable quantity.

The conditional true score criterion variance applicable to Equation 6.13 can be expressed as Equation 6.21:

$$\begin{aligned} \sigma^2[T_Y | T_X] &= \sigma^2[T_Y] (1 - \rho^2[T_Y, T_X]) \\ &= \rho_{tt}Y \sigma^2[Y] \{1 - \rho^2[X, Y] / (\rho_{tt}Y \rho_{tt}X)\} \\ &= \rho_{tt}Y \sigma^2[Y] - \sigma^2[Y] \rho^2[X, Y] / \rho_{tt}X \end{aligned}$$

$$\begin{aligned}
&= \sigma^2[Y](\rho_{ttY} - \rho^2[X,Y]/\rho_{ttX}) \\
&\leq \sigma^2[T_Y | E[T_X | X]] = \sigma^2[T_Y | X] \leq \sigma^2[Y | X] \text{-----} 6.21
\end{aligned}$$

Thus:

$$\sigma[T_Y | T_X] \leq \sigma[T_Y | E[T_X | X_i]] = \sigma[T_Y | X] \leq \sigma[Y | X]$$

Basing selection decision-making on the expected true criterion performance estimated through Equation 6.14 from the regression of  $T_Y$  on the expected true predictor performance conditional on the observed predictor score instead of  $E[Y | X]$  estimated from the regression of  $Y$  on  $X$ , does not affect the number of applicants qualifying for restricted or unrestricted selection.

Correcting the parameters of the regression equation and the standard error of estimate for criterion and predictor unreliability will, however, in the case of on a positive criterion, increase the conditional probability of success for all applicants with  $E[Y] > Y_c$ .

The removal of random measurement error from both the criterion and the predictor will unlikely affect the selection decision if a selection quota is in force and the principle of strict top down selection applies since the corrected conditional probabilities are linearly related to the original, uncorrected probabilities. The same top  $N_v$  applicants will, consequently, still be selected. The correction could only affect restricted selection decision-making if the number of applicants initially fulfilling the selection requirement  $[\alpha_s]$  based on the uncorrected conditional probabilities are less than the number required.

In the case of selection unrestricted by quotas, however, the selection decision will be affected. The preceding results imply that a number of applicants who initially did not succeed in qualifying could in fact now satisfy the entry requirement after the corrections [assuming  $\alpha_s \geq 0.50$ ]. The selection ratio [i.e.  $SR = N_s/N$ ] will consequently increase.

## 6.2.2 Effect Of Statistical Corrections For Restriction Of Range On The Decision Function

### 6.2.2.1 Case 1 [Case B] Restriction Of Range

Case 1 [Case B] restriction of range refers to the situation when both restricted and unrestricted variances are known only for the variable on which selection occurs indirectly/incidentally [Gulliksen, 1950].

The focus of interest in validation research is the regression of Y on X. For the current argument the variable on which selection occurs plays an important role.

If selection occurs directly on the predictor X, but the restricted and unrestricted variances are known only for the criterion Y, then by assumption neither the regression of Y on X nor the criterion variance conditional on X will be affected. Thus, by assumption Equations 6.22 and 6.23 apply. The same notation convention as before is used.

$$\begin{aligned}\beta[Y | X] &= \rho[X, Y](\sigma[Y]/\sigma[X]) \\ &= \rho[x, y](\sigma[y]/\sigma[x]) \\ &= \beta[y | x] \text{-----} 6.22\end{aligned}$$

$$\begin{aligned}\sigma^2[Y | X] &= \sigma^2[Y](1 - \rho^2[X, Y]) \\ &= \sigma^2[y](1 - \rho^2[x, y]) \\ &= \sigma^2[y | x] \text{-----} 6.23\end{aligned}$$

Since Equation 6.22 reflects the assumption that  $\mu[Y | X_i]$  is not altered by explicit selection on X, Equation 6.24 also applies.

$$\begin{aligned}\alpha[Y | X] &= \mu[Y] - \beta[Y | X]\mu[X] \\ &= \mu[y] - \beta[y | x]\mu[x] \\ &= \alpha[y | x] \text{-----} 6.24\end{aligned}$$

Although the magnitude of the correlation is influenced by Case 1 [Case B] selection on X, and although the standard errors of the regression coefficients will increase due to this type of Case 1 [Case B] selection, the relevant decision function [and therefore also the contents of the strategy matrix] will not be affected by Case 1 [Case B] selection on X. Corrections to the parameters of the decision

function to compensate for this type of Case 1 [Case B] restriction of range are consequently not required.

If, in contrast, selection occurs directly on the criterion Y, but the restricted and unrestricted variances are known only for the predictor X, then by assumption neither the regression of X on Y nor the predictor variance conditional on Y will be affected. Thus, by assumption the equivalent of Equations 6.22 - 6.24 would still apply. In a selection context, however, the interest is not in the regression of X on Y but rather in the regression of Y on X. The question consequently arises whether Case 1 [Case B] selection on Y affects the regression of Y on X, and if so, how. Intuitively one would expect that explicit selection on Y should affect the regression of Y on X. Graphical scattergram analysis suggest that the effect of explicit selection on Y on the regression of Y on X should be a flattening of the regression slope [i.e.  $\beta[y|x] < \beta[Y|X]$ ] and a concomitant increase in the intercept [i.e.  $\alpha[y|x] > \alpha[Y|X]$ ]. Equations 6.25 and 6.26 provide analytical confirmation of these inferences by demonstrating that corrections of the parameters of the decision function for the effect of Case 1 [Case B] selection on Y produce an increase in the slope parameter and a decrease in the intercept.

Since by assumption the regression of X on Y is not affected by Case 1 [Case B] selection on Y:

$$\beta[X|Y] = \beta[x|y] \text{-----} 6.25.1$$

Substituting Equation 6.3 in Equation 6.25.1:

$$\rho[X,Y](\sigma[X]/\sigma[Y]) = \rho[x,y](\sigma[x]/\sigma[y]) \text{-----} 6.25.2$$

Multiplying Equation 6.25.2 with  $1/\rho[X,Y]$ :

$$\sigma[X]/\sigma[Y] = (\rho[x,y]/\rho[X,Y])(\sigma[x]/\sigma[y]) \text{-----} 6.25.3$$

Multiplying Equation 6.25.3 with  $1/\sigma[X]$ :

$$1/\sigma[Y] = (\rho[x,y]/\rho[X,Y])(\sigma[x]/\sigma[y]\sigma[X]) \text{-----} 6.25.4$$

Inverting Equation 6.25.4:

$$\sigma[Y] = (\rho[X,Y]\sigma[X]\sigma[y])/(\rho[x,y]\sigma[x]) \text{-----} 6.25.5$$

Substituting Equation 6.3 in Equation 6.25.5:



$$\begin{aligned}
\beta[Y | X] &= \rho[X, Y](\sigma[Y]/\sigma[X]) \\
&= \rho[X, Y](((\rho[X, Y]\sigma[X]\sigma[y])/(\rho[x, y]\sigma[x]))/\sigma[X]) \\
&= (\rho^2[X, Y]\sigma[y])/(\rho[x, y]\sigma[x]) \text{-----} 6.25.6
\end{aligned}$$

However, according to Equation 3.3.5:

$$\begin{aligned}
\rho^2[X, Y] &= 1 - (1 - \rho^2[x, y])(\sigma^2[x]/\sigma^2[X]) \\
&= \{1 - (\sigma^2[x]/\sigma^2[X]) + \rho^2[x, y](\sigma^2[x]/\sigma^2[X])\} \text{-----} 6.25.7
\end{aligned}$$

Substituting Equation 6.25.7 in Equation 6.25.6:

$$\begin{aligned}
\beta[Y | X] &= \{1 - (\sigma^2[x]/\sigma^2[X]) + \rho^2[x, y](\sigma^2[x]/\sigma^2[X])\} \{\sigma[y]/(\rho[x, y]\sigma[x])\} \\
&= \sigma[y]/(\rho[x, y]\sigma[x]) - \sigma[x]\sigma[y]/(\sigma^2[X]\rho[x, y]) + (\rho[x, y]\sigma[x]\sigma[y])/ \sigma^2[X] \\
&= \sigma[y]/(\rho[x, y]\sigma[x]) - (\sigma[x]\sigma[y])/(\rho[x, y]\sigma^2[X]) \{1 - \rho^2[x, y]\} \text{-----} 6.25.8
\end{aligned}$$

Therefore:

$$\begin{aligned}
\beta[Y | X]/\beta[y | x] &= \{\sigma[y]/(\rho[x, y]\sigma[x]) - (\sigma[x]\sigma[y])/(\rho[x, y]\sigma^2[X]) \{1 - \rho^2[x, y]\}\} / \{\rho[x, y](\sigma[y]/\sigma[x])\} \\
&= \{\sigma[y]/(\rho[x, y]\sigma[x]) - (\sigma[x]\sigma[y])/(\rho[x, y]\sigma^2[X]) \{1 - \rho^2[x, y]\}\} \{\sigma[x]/\rho[x, y](\sigma[y])\} \\
&= (1/\rho^2[x, y]) - (\sigma^2[x]/\rho^2[x, y]\sigma^2[X])(1-\rho[x, y]) \\
&= (1/\rho^2[x, y]) - (K^2/\rho^2[x, y])(1-\rho[x, y]) \text{-----} 6.25.9
\end{aligned}$$

Where  $K^2 = \sigma^2[x]/\sigma^2[X]$

Therefore:

$$\beta[Y | X] = \beta[y | x] \{(1/\rho^2[x, y]) - (K^2/\rho^2[x, y])(1-\rho[x, y])\} \text{-----} 6.25$$

According to Equation 6.24:

$$\alpha[X | Y] = \alpha[x | y]$$

Substituting Equation 6.2 in Equation 6.24:

$$\mu[X] - \beta[X | Y]\mu[Y] = \mu[x] - \beta[x | y]\mu[y] \text{-----} 6.26.1$$

Therefore:

$$\beta[X | Y]\mu[Y] = \mu[X] - \mu[x] + \beta[x | y]\mu[y] \text{-----} 6.26.2$$

Therefore, since by assumption,  $\beta[X | Y] = \beta[x | y]$ :

$$\begin{aligned} \mu[Y] &= \beta^{-1}[X | Y](\mu[X] - \mu[x] + \beta[x | y]\mu[y]) \\ &= \beta^{-1}[x | y](\mu[X] - \mu[x] + \beta[x | y]\mu[y]) \\ &= \beta^{-1}[x | y](\mu[X] - \mu[x]) + \mu[y] \text{-----} 6.26.3 \end{aligned}$$

Substituting Equation 6.3 in Equation 6.26.3:

$$\mu[Y] = (\rho[x,y]\sigma[x]/\sigma[y])^{-1}(\mu[X] - \mu[x]) + \mu[y] \text{-----} 6.26.4$$

Substituting Equation 6.26.4 in Equation 6.2:

$$\begin{aligned} \alpha[Y | X] &= \mu[Y] - \beta[Y | X]\mu[X] \\ &= (\rho[x,y]\sigma[x]/\sigma[y])^{-1}(\mu[X] - \mu[x]) + \mu[y] - \beta[Y | X]\mu[X] \text{-----} 6.26 \end{aligned}$$

With  $\beta[Y | X]$  estimated by Equation 6.25

The expression for the regression of Y on X, corrected for Case 1 [Case B] selection on Y, is shown as Equation 6.27.

$$E[Y | X] = \alpha[Y | X] + \beta[Y | X]X_i \text{-----} 6.27.1$$

Substituting Equations 6.25 and 6.26 into Equation 6.27.1:

$$\begin{aligned} E[Y | X] &= (\rho[x,y]\sigma[x]/\sigma[y])^{-1}(\mu[X] - \mu[x]) + \mu[y] - \beta[Y | X]\mu[X] + \beta[Y | X]X_i \\ &= (\rho[x,y]\sigma[x]/\sigma[y])^{-1}(\mu[X] - \mu[x]) + \mu[y] + \beta[Y | X]\{X_i - \mu[X]\} \\ &= (\rho[x,y]\sigma[x]/\sigma[y])^{-1}(\mu[X] - \mu[x]) + \mu[y] + \{\sigma[y]/(\rho[x,y]\sigma[x]) - \\ &\quad (\sigma[x]\sigma[y])/(\rho[x,y]\sigma^2[X])(1 - \rho^2[x,y])\}\{X_i - \mu[X]\} \text{-----} 6.27 \end{aligned}$$

The change in the regression parameters implies that the regression equation derived from the selected applicant group and the regression equation that would have been obtained from the unrestricted applicant group will intersect at some point  $X_s$  on the X-axis. Consequently, for all  $X_i < X_s$   $E[y | x_i]$  will overestimate  $E[Y | X_i]$  while for all  $X_i > X_s$   $E[y | x_i]$  will underestimate  $E[Y | X_i]$ . Correcting the parameters of the decision function for Case 1 [Case B] selection on Y, consequently will have the

effect of elevating the expected criterion performance for all  $X_i > X_s$ , while depressing it for all  $X_i < X_s$ .

The correction of the decision function for Case 1 [Case B] selection on Y will not affect the applicants selected if selection is restricted by quotas and strict top down selection applies. Although  $E[Y | X_i] \neq E[y | x_i]$ , if  $E[y | x_i] > E[y | x_j]$  then  $E[Y | X_i] > E[Y | X_j]$ . The same  $N_v$  applicants would therefore still be selected.

In the case of selection unrestricted by quotas, however, the use of the corrected decision function will affect the selection decision-making. The nature of the effect will depend on the position of  $X_c$  [derived from  $Y_c$  via the decision function calculated on the selected applicant group] relative to  $X_s$ . If  $X_s < X_c$  the number of applicants satisfying the entry requirement [i.e.  $E[Y | X_i] > Y_c$ ] will increase. If  $X_s > X_c$  [unlikely since it would imply a very low  $Y_c$  on a positive criterion], the number of applicants qualifying for selection will decrease.

Case 1 [Case B] selection on X does not affect the standard error of estimate  $\sigma[Y | X]$  [Equation 6.23]. Case 1 [Case B] selection on Y, however, does affect the standard error of estimate  $\sigma[Y | X]$ . Equation 6.28 provides a correction to the standard error of estimate  $\sigma[y | x]$  for Case 1 [Case B] selection on Y.

Since by assumption the regression of X on Y is not affected by Case 1 [Case B] selection on Y:

$$\beta[x | y] = \beta[X | Y] \text{-----} 6.28.1$$

Substituting Equation 6.3 into Equation 6.28.1:

$$\rho[x,y](\sigma[x]/\sigma[y]) = \rho[X,Y](\sigma[X]/\sigma[Y]) \text{-----} 6.28.2$$

Multiplying Equation 6.28.2 with  $\sigma[Y]\sigma[y]/\rho[x,y]\sigma[x]$ :

$$\sigma[Y] = (\rho[X,Y]\sigma[X]\sigma[y])/(\rho[x,y]\sigma[x]) \text{-----} 6.28.3$$

Substituting Equation 6.28.3 into Equation 6.5:

$$\begin{aligned} \sigma^2[Y | X] &= \sigma^2[Y](1 - \rho^2[X,Y]) \\ &= \{(\rho^2[X,Y]\sigma^2[X]\sigma^2[y])/(\rho^2[x,y]\sigma^2[x])\}(1 - \rho^2[X,Y]) \text{-----} 6.28.4 \end{aligned}$$

Multiplying the terms in Equation 6.28.4:

$$\sigma^2[Y|X] = \{(\rho^2[X,Y]\sigma^2[X]\sigma^2[Y])/(\rho^2[x,y]\sigma^2[x])\} - \{(\rho^4[X,Y]\sigma^2[X]\sigma^2[Y])/(\rho^2[x,y]\sigma^2[x])\} \text{-----} 6.28.5$$

Rewrite Equation 6.28.5 with  $K = \sigma[x]/\sigma[X]$ :

$$\begin{aligned} \sigma^2[Y|X] &= \{(\rho^2[X,Y]\sigma^2[Y])/(\rho^2[x,y]K^2)\} - \{(\rho^4[X,Y]\sigma^2[Y])/(\rho^2[x,y]K^2)\} \\ &= \sigma[Y]\{\rho^2[X,Y]/(\rho^2[x,y]K^2) - \rho^4[X,Y]/(\rho^2[x,y]K^2)\} \\ &= \sigma[Y]\{(\rho^2[X,Y] - \rho^4[X,Y])/(\rho^2[x,y]K^2)\} \\ &= \sigma[Y]\{\rho^2[X,Y](1 - \rho^2[X,Y])/(\rho^2[x,y]K^2)\} \text{-----} 6.28.6 \end{aligned}$$

However, the validity coefficient for the unrestricted applicant group is given by Equation 3.3:

$$\begin{aligned} \rho[X,Y] &= \sqrt{\{1 - (1 - \rho^2[x,y])(\sigma^2[x]/\sigma^2[X])\}} \\ &= \sqrt{\{1 - K^2(1 - \rho^2[x,y])\}} \end{aligned}$$

Squaring Equation 3.3 and substituting in Equation 6.28.6:

$$\begin{aligned} \sigma^2[Y|X] &= \sigma[Y]\{\{1 - K^2(1 - \rho^2[x,y])\} (1 - \{1 - K^2(1 - \rho^2[x,y])\})/(\rho^2[x,y]K^2)\} \\ &= \sigma[Y]\{\{1 - K^2(1 - \rho^2[x,y])\} (K^2(1 - \rho^2[x,y]))/(\rho^2[x,y]K^2)\} \\ &= \sigma[Y]\{(K^2(1 - \rho^2[x,y])) - (K^4(1 - \rho^2[x,y])^2)/(\rho^2[x,y]K^2)\} \\ &= \sigma[Y]\{(1 - \rho^2[x,y]) - (K^2(1 - \rho^2[x,y])^2)/\rho^2[x,y]\} \\ &= \sigma[Y]\{\{(1 - \rho^2[x,y])\}\{1 - (K^2(1 - \rho^2[x,y])^2)/\rho^2[x,y]\}\} \text{-----} 6.28 \\ &\geq \sigma^2[y|x] \end{aligned}$$

Analysis of Equation 6.28 indicates that Case 1 [Case B] selection on Y produces an decrease in the standard error of estimate [i.e.  $\sigma[y|x] < \sigma[Y|X]$ ]. Correcting  $\sigma[y|x]$  for Case 1 [Case B] selection on Y via Equation 6.28 thus returns a corrected value  $\sigma[Y|X]$  greater than the obtained standard error of estimate. The difference between the corrected and uncorrected standard errors of estimate seems to be negatively related to the observed correlation  $\rho[x,y]$ .

An analysis of the effect of corrections for Case 1 [Case B] selection on Y, on selection decision making based on the conditional probability of success, should focus on the relative position of  $Y_c$  in the conditional criterion distribution. Since the critical criterion cutoff value,  $Y_c$ , is defined on a scale unaffected by [systematic and/or random] measurement error, the relative position of  $Y_c$  in the conditional criterion distribution depends on the reaction of both the expected criterion performance and the standard error of estimate to corrections for Case 1 [Case B] selection on Y. A factor complicating the issue, however, is the fact that the effect of the change in expected criterion performance on the conditional probability tends to oppose the effect of the change in the standard

error of estimate. The change in the expected criterion performance exerts an upward pressure on the conditional probability for all  $X_i > X_s$  and a downward pressure for all  $X_i < X_s$ . The change in the standard error of estimate, in contrast, produces the opposite effect. The reaction of the conditional probability of success to the aforementioned corrections for Case 1 [Case B] selection on Y, consequently, depends on the which one of the two processes dominate. However, no single answer to this question exists. Since the magnitude of the change in the expected criterion performance depends on  $X_i$ , but the change in the standard error of estimate is constant over all values of  $X_i$ , the question on which one of the processes would dominate would also depend on  $X_i$ .

Logic would suggest two points on the X-axis [ $X_{hi}$  &  $X_{lo}$ ], positioned such on both sides of  $X_s$ , that  $Z[Y_c; E[y|x=X_{hi}]] = Z[Y_c; E[Y|X=X_{hi}]]$  and  $Z[Y_c; E[y|x=X_{lo}]] = Z[Y_c; E[Y|X=X_{lo}]]$ . The conditional probability of success associated with applicants falling on  $X_{hi}$  or  $X_{lo}$  would therefore not be affected by the preceding corrections for Case 1 [Case B] selection on Y. However, for all  $X_i > X_{hi}$ , and all  $X_i < X_{lo}$ , the effect of the change in expected criterion performance would dominate. Consequently the conditional probability of success associated with those applicants falling above  $X_{hi}$  will increase [assuming  $Y_c > \mu[Y]$ ], while the conditional probability of success associated with those applicants falling below  $X_{lo}$  will decrease. Furthermore, for those applicants located between these two cutoff points [i.e.  $X_{lo} < X_i < X_{hi}$ ], the effect of the change in the standard error of estimate should dominate and consequently the conditional probability should increase.

The application of Equations 6.27 and 6.28 will [probably] not affect selection decision-making based on the conditional probability of success, if selection quotas restrict such selection.

The application of Equations 6.27 and 6.28 will, however, affect unrestricted selection decision-making based on the conditional probability of success. The selection ratio should increase since the critical acceptance probability  $\alpha_s$  is constant while the conditional probability of success associated with those applicants falling above  $X_{hi}$  increases. The number of applicants qualifying for selection therefore increases.

The probability of Case 1 [Case B] selection on X or Y occurring in the context of selection validation research appears to be rather remote, although not altogether impossible. The chance to apply the foregoing correction formula would therefore seldom arise.

### 6.2.2.2 Case 2 [Case A] Restriction Of Range

Case 2 [Case A] restriction of range refers to the situation where restricted and unrestricted variances are known only for the variable on which selection occurred. Thus either selection occurs directly on the predictor and both  $\sigma^2[x]$  and  $\sigma^2[X]$  are known, or selection occurs on the criterion and both  $\sigma^2[y]$  and  $\sigma^2[Y]$  are known [Gulliksen, 1950].

The focus of interest in validation research is the regression of Y on X. If Case 2 [Case A] selection occurs directly on the predictor X, then by assumption, neither the regression of Y on X nor the criterion variance conditional on X will be affected. Thus, by assumption Equations 6.22 - 6.24 apply. No corrections to the parameters of the regression equation or the standard error of estimate are therefore required. The regression of X on Y would be affected, but since it is of no real interest in selection validation research, no justification seems to exist to explore it further.

If Case 2 [Case A] selection occurs directly on the criterion Y, then by assumption, neither the regression of X on Y nor the predictor variance conditional on Y will be affected. The regression of Y on X and the criterion variance conditional on X would be affected. The probability of Case 2 [Case A] selection on Y actually occurring in the context of selection validation research, however, seems for all practical purposes to be zero. Although appropriate correction formula therefore should exist, there once more seems to be no practical justification to try and uncover their formulation.

### 6.2.2.3 Case C Restriction Of Range

Case C restriction of range refers to a situation where selection occurs on a variable other than the predictor and criterion variables being correlated. Both the predictor and criterion variables are therefore only subject to incidental selection to the extent that they are correlated with the third variable Z on which explicit selection occurs. Case 3[i] [Case C[i]] restriction of range refers to a situation where the restricted and unrestricted variances are only known for the explicit selection variable. Case 3[ii] [Case C[ii]] restriction of range refers to a situation where the restricted and unrestricted variances are known only for one of the incidental selection variables [Gulliksen, 1950; Thorndike, 1949; Thorndike, 1982].

Case C selection correction formula for the correlation coefficient are based on the assumptions that [Gulliksen, 1950]:

$$\triangleright \beta[y|z] = \beta[Y|Z] \text{ and } \beta[x|z] = \beta[X|Z];$$



- $\sigma[y|z] = \sigma[Y|Z]$  and  $\sigma[x|z] = \sigma[X|Z]$ ;
- $\rho[x,y|z] = \rho[X,Y|Z]$

In terms of these formal assumptions the coefficients characterising the regression of X or Y on Z are not affected by Case C selection. No corrections to the slope and intercept parameters are therefore required. The regression of X or Y on Z is, however, not of any significant interest in selection validation research. The focus of interest in validation research is the regression of Y on X. The coefficients in the regression of Y on X are, however, affected by Case C selection on Z. Neither the slope and intercept parameters of the regression of Y on X, nor the variance of the [homoscedastic] conditional criterion distributions, derived on the restricted population therefore would be applicable to the unrestricted population.

Of the cases considered Case C selection probably most closely resembles the process actually operating in selection validation research. Sufficient practical justification therefore seems to exist to try and uncover appropriate correction formula.

Equation 6.29 provides an expression for the slope of the regression of Y on X for the unselected population. Equation 6.29 assumes Case 3[i] [Case C[i]] selection on Z.

The slope of the regression of the criterion on the predictor is given by Equation 6.3:

$$\beta[Y|X] = \rho[X,Y](\sigma[Y]/\sigma[X]) \text{-----} 6.29.1$$

However, the terms  $\rho[X,Y]$ ,  $\sigma[Y]$  and  $\sigma[X]$  in Equation 6.29.1 are all unknown under Case 3[i] [Case C[i]] conditions

Furthermore, by assumption the regression of Y on Z is not affected by Case 3[i] [Case C[i]] selection on Z:

$$\beta[Y|Z] = \beta[y|z] \text{-----} 6.29.2$$

Substituting the appropriate equivalent of Equation 6.3 into Equation 6.29.2:

$$\rho[Y,Z](\sigma[Y]/\sigma[Z]) = \rho[y,z](\sigma[y]/\sigma[z]) \text{-----} 6.29.3$$

Isolating the term  $\rho[Y,Z]$  in Equation 6.29.3:

$$\rho[Y,Z] = \rho[y,z](\sigma[y]\sigma[Z]/(\sigma[z]\sigma[Y])) \text{-----} 6.29.4$$

Furthermore, by assumption the criterion variance conditional on Z, is not affected by Case 3[i] [Case C[i]] selection on Z:

$$\sigma[y|z] = \sigma[Y|Z] \text{-----} 6.29.5$$

Squaring both sides of Equation 6.29.5 and substituting the equivalent of Equation 6.5 into Equation 6.29.5:

$$\sigma^2[y](1 - \rho^2[y,z]) = \sigma^2[Y](1 - \rho^2[Y,Z]) \text{-----} 6.29.6$$

Substituting Equation 6.29.4 into Equation 6.29.6 and simplifying:

$$\begin{aligned} \sigma^2[y](1 - \rho^2[y,z]) &= \sigma^2[Y](1 - \{\rho^2[y,z]((\sigma^2[y]\sigma^2[Z])/(\sigma^2[z]\sigma^2[Y]))\}) \\ &= \sigma^2[Y] - \rho^2[y,z]\sigma^2[y](\sigma^2[Z]/\sigma^2[z]) \\ &= \sigma^2[Y] - \rho^2[y,z]\sigma^2[y]K^2 \text{-----} 6.29.7 \end{aligned}$$

where,  $K = \sigma[Z]/\sigma[z]$

Isolating  $\sigma^2[Y]$  in Equation 6.29.7:

$$\begin{aligned} \sigma^2[Y] &= \sigma^2[y](1 - \rho^2[y,z]) + \rho^2[y,z]\sigma^2[y]K^2 \\ &= \sigma^2[y](1 - \rho^2[y,z] + \rho^2[y,z]K^2) \text{-----} 6.29.8 \end{aligned}$$

By assumption the predictor variance conditional on Z is not affected by Case 3[i] [Case C[i]] selection on Z:

$$\sigma[x|z] = \sigma[X|Z] \text{-----} 6.29.9$$

Squaring both sides of Equation 6.29.9, substituting the equivalent of Equation 6.5 into Equation 6.29.9 and applying the equivalent logic as before:

$$\sigma^2[X] = \sigma^2[x](1 - \rho^2[x,z] + \rho^2[x,z]K^2) \text{-----} 6.29.10$$

Substituting Equation 3.5 and Equations 6.29.8 and 6.29.9 into Equation 6.29.1:

$$\begin{aligned} \beta[Y|X] &= \rho[X,Y](\sigma[Y]/\sigma[X]) \\ &= \{\{\rho[x,y]-\rho[x,z]\rho[y,z]\} + \{\rho[x,z]\rho[y,z]K^2\}\} / \{\sqrt{1 - \rho^2[x,z] + \rho^2[x,z]K^2}\} \sqrt{1 - \rho^2[y,z] + \rho^2[y,z]K^2}\} \{\sigma[y] \sqrt{1 - \rho^2[y,z] + \rho^2[y,z]K^2}\} \end{aligned}$$

$$\frac{\rho^2[y,z]K^2)}{\{\sigma[x]\sqrt{(1 - \rho^2[x,z] + \rho^2[x,z]K^2)}\}} \\ = \frac{\{\{\rho[x,y]-\rho[x,z]\rho[y,z]\} + \{\rho[x,z]\rho[y,z]K^2\}\}}{\{(1 - \rho[x,z] + \rho^2[x,z]K^2)\}\{\sigma[y]/\sigma[x]\}} \text{-----} 6.29$$

Linn [1983, p. 6] derived an expression for  $\beta[Y|X]$ , under Case 3[i] [Case C[i]] selection conditions, identical to Equation 6.29

Equation 6.30 provides an expression for the intercept of the regression of Y on X for the unselected population derived from  $\alpha[y|x]$  and Case 3[i] [Case C[i]] selection.

The Y-axis intercept of the regression of Y on X is given by Equation 6.2:

$$\alpha[Y|X] = \mu[Y] - \beta[Y|X]\mu[X] \text{-----} 6.30.1$$

However,  $\mu[Y]$  and  $\mu[X]$  are not known under Case 3[i] [Case C[i]] selection conditions.

By assumption the intercept of the regression of Y on Z is not affected by Case 3[i] [Case C[i]] selection on Z:

$$\alpha[Y|Z] = \alpha[y|z] \text{-----} 6.30.2$$

Substituting the equivalent of Equation 6.2 into Equation 6.30.2:

$$\mu[Y] - \beta[Y|Z]\mu[Z] = \mu[y] - \beta[y|z]\mu[z] \text{-----} 6.30.3$$

Isolating the mean of the unrestricted criterion distribution in Equation 6.30.3:

$$\mu[Y] = \mu[y] - \beta[y|z]\mu[z] + \beta[Y|Z]\mu[Z] \text{-----} 6.30.4$$

Since the regression of Y on Z is not affected by Case 3[i] [Case C[i]] selection on Z, Equation 6.30.4 can be simplified:

$$\mu[Y] = \mu[y] + \beta[y|z](\mu[Z] - \mu[z]) \text{-----} 6.30.5$$

Substituting the expression for the slope of the regression of y on x into Equation 6.30.5:

$$\mu[Y] = \mu[y] + \rho[y,z](\sigma[y]/\sigma[z])(\mu[Z] - \mu[z]) \text{-----} 6.30.6$$

Similarly, since by assumption the Y-axis intercept of the regression of X on Y is also not affected by Case 3[i] [Case C[i]] selection on Z:

$$\alpha[X|Z] = \alpha[x|z] \text{-----} 6.30.7$$

Applying the same argument in terms of which Equation 6.30.6 was derived from Equation 6.30.2 to Equation 6.30.7:

$$\begin{aligned} \mu[X] &= \mu[x] + \beta[x|z](\mu[Z] - \mu[z]) \\ &= \mu[x] + \rho[x,z](\sigma[x]/\sigma[z])(\mu[Z] - \mu[z]) \text{-----} 6.30.8 \end{aligned}$$

Substituting Equations 6.30.6 and 6.30.8 in Equation 6.30.1:

$$\begin{aligned} \alpha[Y|X] &= \mu[Y] - \beta[Y|X]\mu[X] \\ &= \{\mu[y] + \rho[y,z](\sigma[y]/\sigma[z])(\mu[Z] - \mu[z])\} - \\ &\quad \beta[Y|X]\{\mu[x] + \rho[x,z](\sigma[x]/\sigma[z])(\mu[Z] - \mu[z])\} \text{-----} 6.30 \end{aligned}$$

Where:

$\beta[Y|X]$  is derived via Equation 6.29.

Graphical analysis of Equations 6.29 and 6.30 suggest that these corrections for Case 3[i] [Case C[i]] selection on Z can have either an accelerative or a restraining effect on the slope parameter. The reaction of the ratio  $\Delta = \beta[Y|X]/\beta[y|x]$  [delta] to changes in K,  $\rho[x,y]$ ,  $\rho[x,z]$  and  $\rho[y,z]$  [the latter three were assumed positive] is graphically depicted in Figures 6.2 - 6.9 For the purpose of the computational solution the predictor and criterion were assumed to be standardised [0; 1] variables. Figures 6.2 - 6.9 seem to suggest that Equations 6.29 and 6.30 generally have the effect of increasing the slope of the regression of Y on X relative to the slope of the regression of y on x and, concomitantly, decreasing the intercept. There are, however, conditions where this is clearly not the case. Linn [1983, p. 6] seems to arrive at an equivalent conclusion when he reports:

"A little work with Equations 3 and 4 will show that the typical effect of selection is to flatten the slope and produce a concomitant increase in the intercept. This statement assumes positively related variables of reasonable magnitude. It is possible, as Levin [1972] demonstrated for the correlation between X and Y, for the selection to increase the slope. But this result occurs only for combinations of correlations that are unlikely to be encountered in practice [for example, when the selection variable is correlated with X but not with Y]."

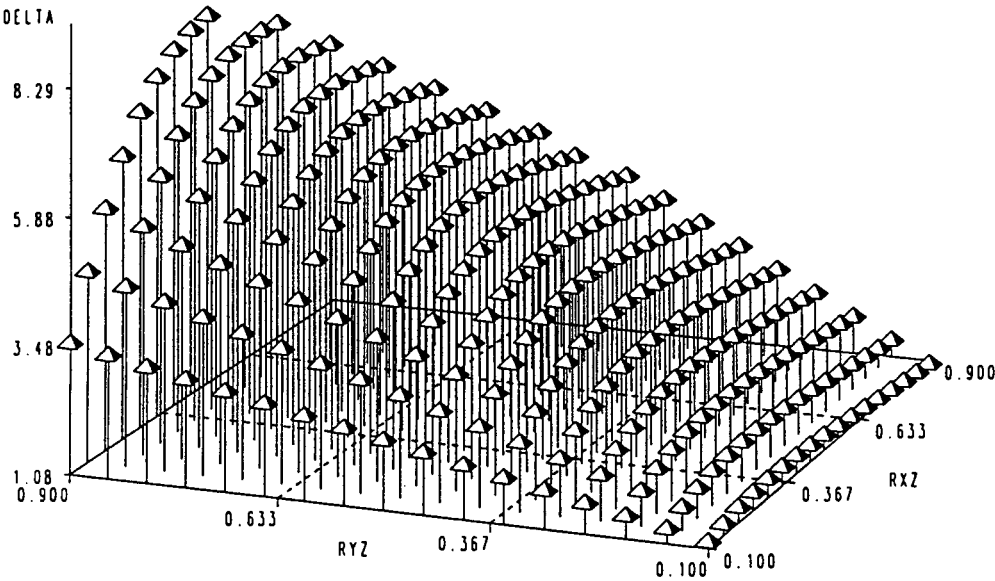


Figure 6.2: The ratio  $\Delta = \beta[Y|X]/\beta[y|x][\text{delta}]$  as a function of  $\rho[x,z]$  and  $\rho[y,z]$  for  $\rho[x,y]$  fixed at 0.10 and  $K$  fixed at 2

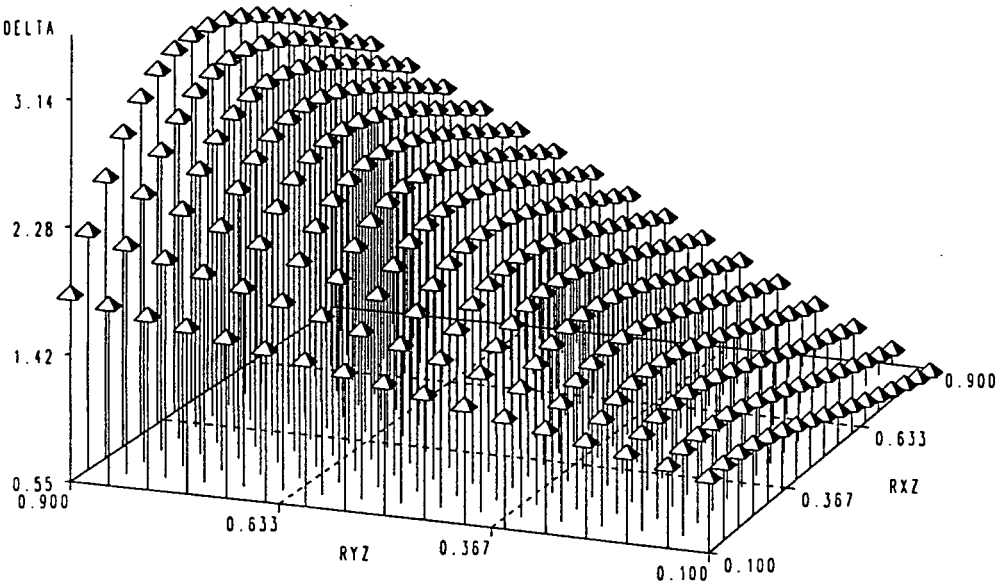


Figure 6.3: The ratio  $\Delta = \beta[Y|X]/\beta[y|x][\text{delta}]$  as a function of  $\rho[x,z]$  and  $\rho[y,z]$  for  $\rho[x,y]$  fixed at 0.30 and  $K$  fixed at 2

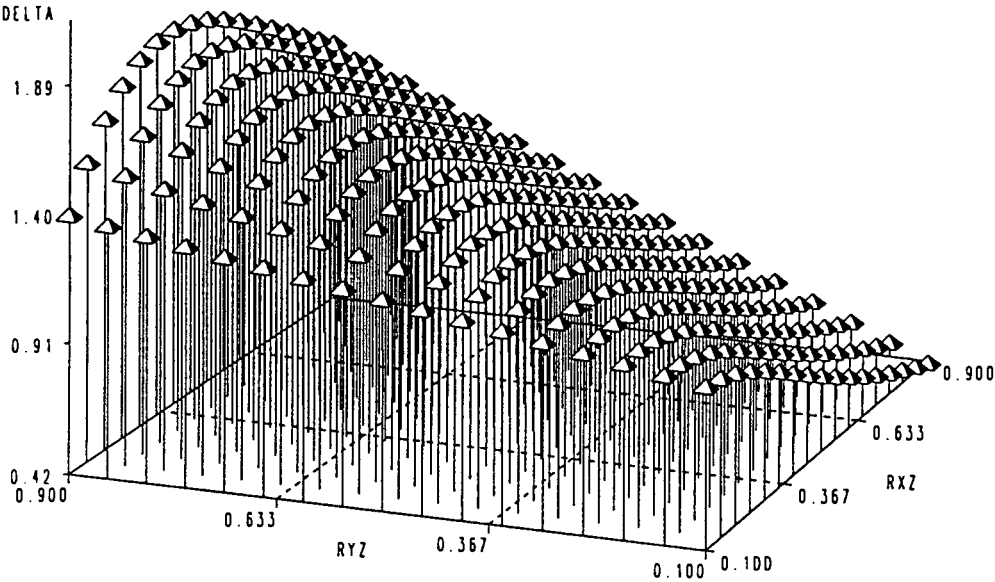


Figure 6.4: The ratio  $\Delta = \beta[Y|X]/\beta[y|x]$  [delta] as a function of  $\rho[x,z]$  and  $\rho[y,z]$  for  $\rho[x,y]$  fixed at 0.60 and K fixed at 2

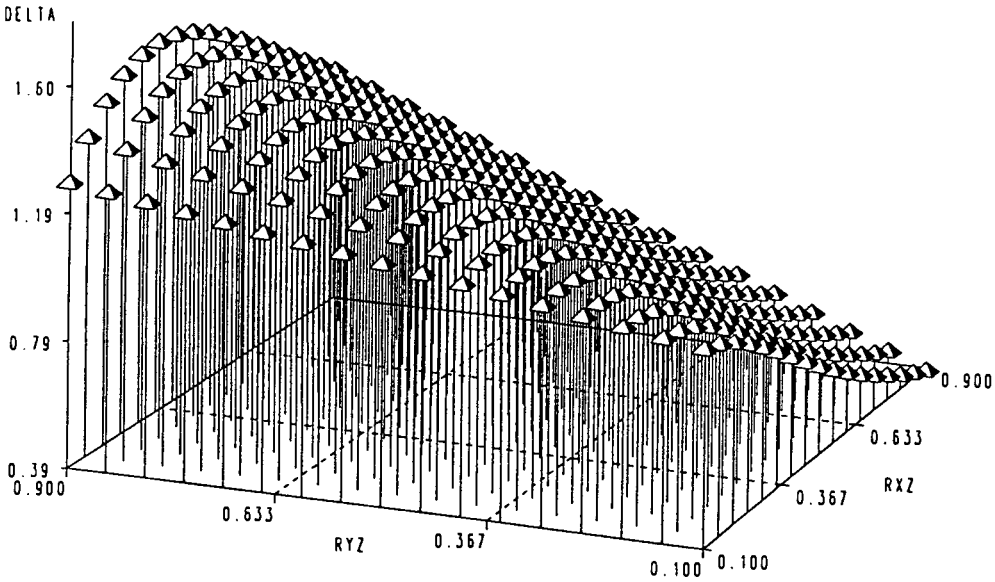


Figure 6.5: The ratio  $\Delta = \beta[Y|X]/\beta[y|x]$  [delta] as a function of  $\rho[x,z]$  and  $\rho[y,z]$  for  $\rho[x,y]$  fixed at 0.80 and K fixed at 2



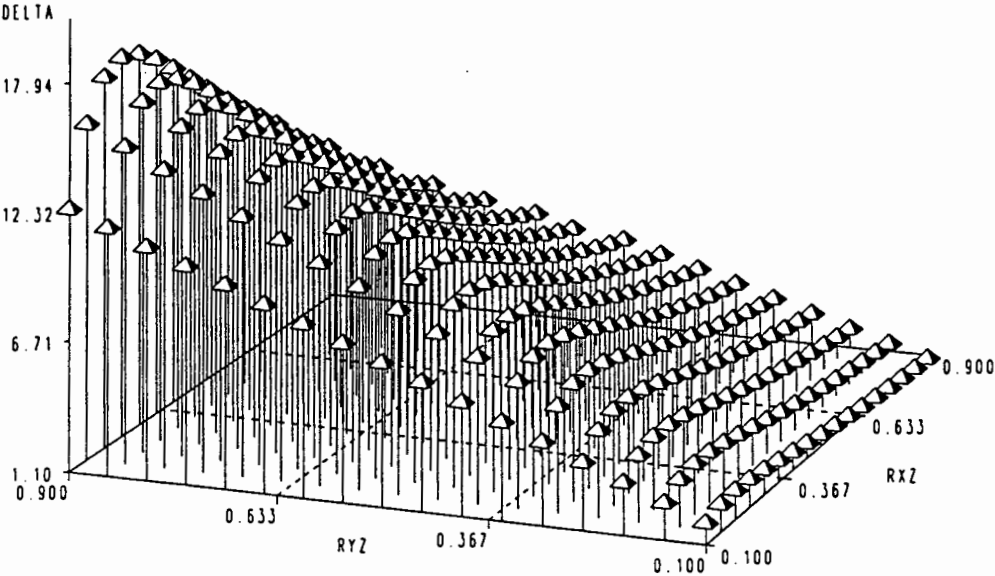


Figure 6.6: The ratio  $\Delta = \beta[Y|X]/\beta[y|x]$  [delta] as a function of  $\rho[x,z]$  and  $\rho[y,z]$  for  $\rho[x,y]$  fixed at 0.10 and  $K$  fixed at 4.

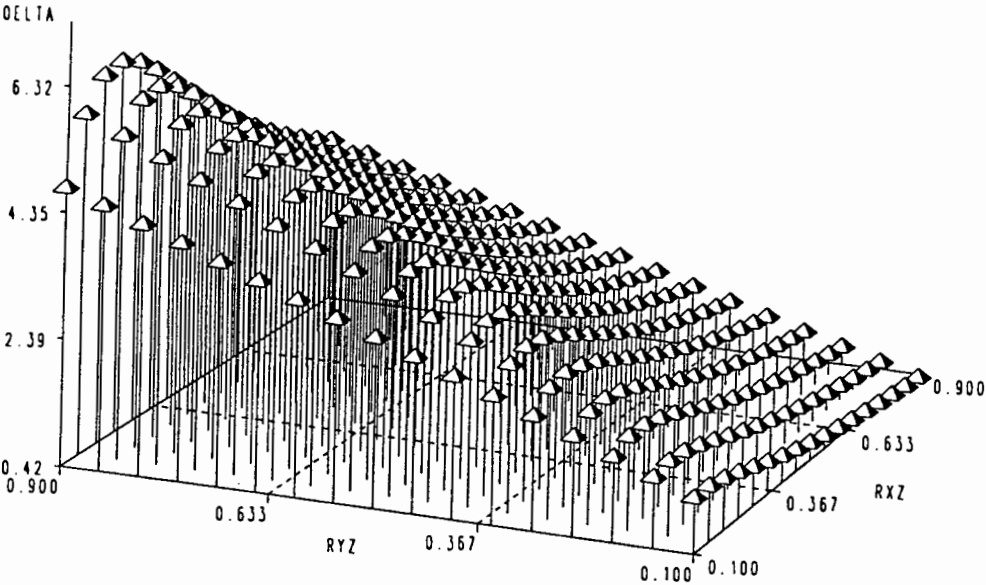


Figure 6.7: The ratio  $\Delta = \beta[Y|X]/\beta[y|x]$  [delta] as a function of  $\rho[x,z]$  and  $\rho[y,z]$  for  $\rho[x,y]$  fixed at 0.30 and  $K$  fixed at 4.

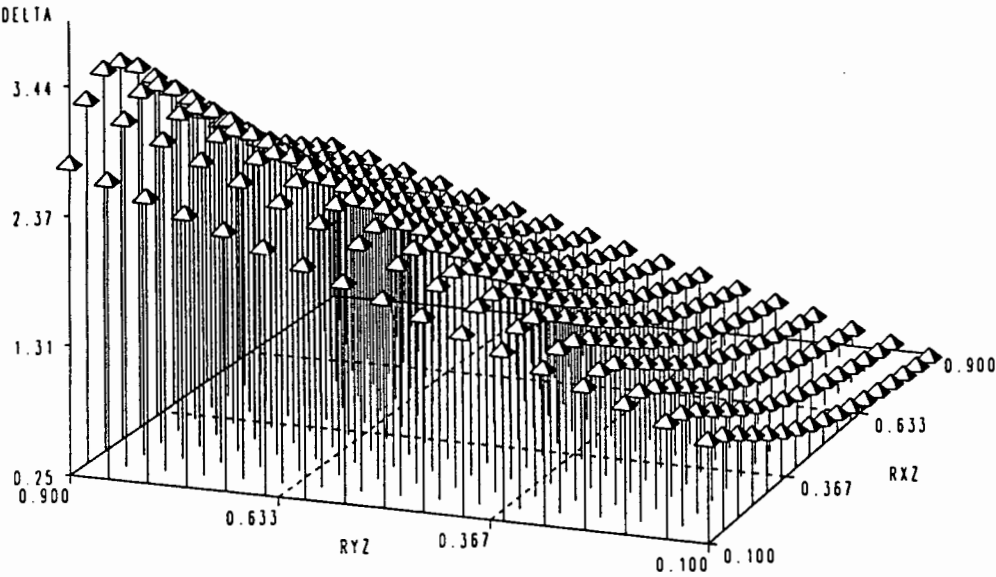


Figure 6.8: The ratio  $\Delta = \beta[Y|X]/\beta[y|x]$  [delta] as a function of  $\rho[x,z]$  and  $\rho[y,z]$  for  $\rho[x,y]$  fixed at 0.60 and K fixed at 4

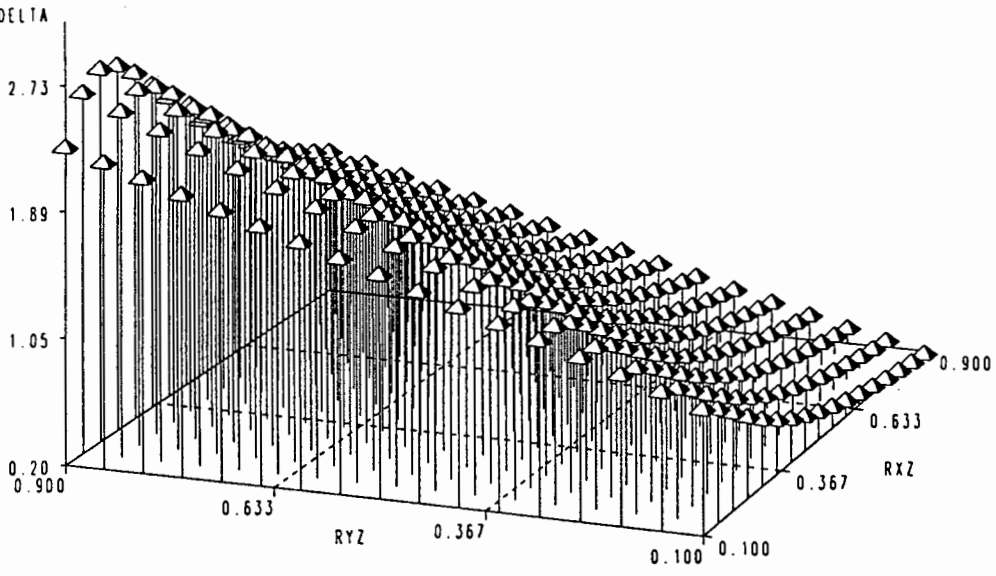


Figure 6.9: The ratio  $\Delta = \beta[Y|X]/\beta[y|x]$  [delta] as a function of  $\rho[x,z]$  and  $\rho[y,z]$  for  $\rho[x,y]$  fixed at 0.80 and K fixed at 4

The regression of  $y$  on  $x$  and the estimated/corrected regression of  $Y$  on  $X$  will intersect at some point  $X_s$  on the  $X$ -axis. The magnitude and direction of the change in the expected criterion performance associated with applicants attributable to the application of Equations 6.29 and 6.30 will depend on the position of  $X_i$  relative to  $X_s$  and the magnitude of  $\beta[Y|X]$  relative to  $\beta[y|x]$ . Table 6.1 presents a summary of the effect of Equations 6.29 and 6.30 on the expected criterion performance associated with applicants.

Table 6.1: The effect of Equations 6.29 and 6.30 on the expected criterion performance associated with applicants

	$\beta[Y X] < \beta[y x]$	$\beta[Y X] > \beta[y x]$
$X_i < X_s$	$E[Y X] > E[y x]$	$E[Y X] < E[y x]$
$X_i > X_s$	$E[Y X] < E[y x]$	$E[Y X] > E[y x]$

Correction for Case 3[i] [Case C[i]] selection on  $Z$  will [probably] not affect selection decision-making based on expected criterion performance when selection quotas limit the number of applicants required. Correcting the bias in the parameters of the restricted regression function would not alter the initial rank-order based on  $E[y|x]$ . The same  $N_s$  applicants will thus still be selected as long as the principle of strict top down selection is adhered to.

In the case of unrestricted selection, however, corrections for Case C selection on  $Z$  probably would affect selection decision-making. Since  $Y_c$  is assumed constant, the change in the expected criterion performance attributable to Equations 6.29 and 6.30 could move a number of applicants across the critical criterion cutoff. The direction and magnitude of the migration would depend on the position of the predictor cutoff derived from  $Y_c$  under the restricted condition  $[x_c]$  relative to  $X_s$  and the magnitude of  $\beta[Y|X]$  relative to  $\beta[y|x]$ . Table 6.2 presents a summary of the anticipated effect of Equations 6.29 and 6.30 on the number of applicants fulfilling the entry requirement for selection [i.e.  $E[Y|X] \geq Y_c$ ].

Table 6.2: The effect of Equations 6.29 and 6.30 on the number of applicants qualifying for selection  $[N_s]$

	$\beta[Y X] < \beta[y x]$	$\beta[Y X] > \beta[y x]$
$x_c < X_s$	$X_c < x_c \therefore N_s \text{ increases}$	$X_c > x_c \therefore N_s \text{ decreases}$
$x_c > X_s$	$X_c > x_c \therefore N_s \text{ decreases}$	$X_c < x_c \therefore N_s \text{ increases}$

The effect of corrections for Case 3[i] [Case C[i]] selection on  $Z$  on the standard error of estimate, is examined through Equation 6.31.

The standard error of estimate for the unrestricted applicant population is given by the square root of Equation 6.5:

$$\sigma[Y|X] = \sigma[Y]\sqrt{(1 - \rho^2[X,Y])} \text{-----} 6.31.1$$

However,  $\rho[X,Y]$  and  $\sigma[Y]$  are both unknown under Case 3[i] [Case C[i]] conditions

By assumption the regression of Y on Z is not effected by Case 3[i] [Case C[i]] selection on Z:

$$\beta[Y|Z] = \beta[y|z] \text{-----} 6.31.2$$

Substituting the appropriate equivalent of Equation 6.3 into Equation 6.31.2:

$$\rho[Y,Z](\sigma[Y]/\sigma[Z]) = \rho[y,z](\sigma[y]/\sigma[z]) \text{-----} 6.31.3$$

Isolating the term  $\rho[Y,Z]$  in Equation 6.31.3:

$$\rho[Y,Z] = \rho[y,z]((\sigma[y]\sigma[Z])/(\sigma[z]\sigma[Y])) \text{-----} 6.31.4$$

Furthermore, by assumption the criterion variance conditional on Z is also not affected by Case 3[i] [Case C[i]] selection on Z:

$$\sigma[y|z] = \sigma[Y|Z] \text{-----} 6.31.5$$

Substituting the appropriate equivalent of Equation 6.5 in Equation 6.31.5:

$$\sigma^2[y](1 - \rho^2[y,z]) = \sigma^2[Y](1 - \rho^2[Y,Z]) \text{-----} 6.31.6$$

Substituting Equation 6.31.4 in Equation 6.31.6:

$$\begin{aligned} \sigma^2[y](1 - \rho^2[y,z]) &= \sigma^2[Y](1 - \{\rho^2[y,z](\sigma^2[y]\sigma^2[Z])/(\sigma^2[z]\sigma^2[Y])\}) \\ &= \sigma^2[Y] - \rho^2[y,z]\sigma^2[y](\sigma^2[Z]/\sigma^2[z]) \\ &= \sigma^2[Y] - \rho^2[y,z]\sigma^2[y]K^2 \text{-----} 6.31.7 \end{aligned}$$

Where,  $K = \sigma[Z]/\sigma[z]$

Isolating the unrestricted criterion variance term in Equation 6.31.7:

$$\begin{aligned}\sigma^2[Y] &= \sigma^2[y](1 - \rho^2[y,z]) + \rho^2[y,z]\sigma^2[y]K^2 \\ &= \sigma^2[y](1 - \rho^2[y,z] + \rho^2[y,z]K^2) \text{-----} 6.31.8\end{aligned}$$

Substituting Equation 6.31.8 and the square of Equation 3.5 in Equation 6.31.1:

$$\begin{aligned}\sigma^2[Y|X] &= \sigma^2[Y](1 - \rho^2[X,Y]) \\ &= \sigma^2[y](1 - \rho^2[y,z] + \rho^2[y,z]K^2)\{1 - \{(\rho[x,y]-\rho[x,z]\rho[y,z]) + \\ &\quad (\rho[x,z]\rho[y,z]K^2)\}/\{\sqrt{(1 - \rho^2[x,z] + \rho^2[x,z]K^2)}\sqrt{(1 - \rho^2[y,z] + \rho^2[y,z]K^2)}\}\}^2\} \\ &= \sigma^2[y](1 - \rho^2[y,z] + \rho^2[y,z]K^2) - \{(\rho[x,y]-\rho[x,z]\rho[y,z]) + \\ &\quad (\rho[x,z]\rho[y,z]K^2)^2\}/(1 - \rho^2[x,z] + \rho^2[x,z]K^2) \text{-----} 6.31.9\end{aligned}$$

Thus:

$$\begin{aligned}\sigma[Y|X] &= \sigma[y]\sqrt{(1 - \rho^2[y,z] + \rho^2[y,z]K^2)} - (\rho[x,y]-\rho[x,z]\rho[y,z]) + \\ &\quad (\rho[x,z]\rho[y,z]K^2)/\sqrt{(1 - \rho^2[x,z] + \rho^2[x,z]K^2)} \text{-----} 6.31\end{aligned}$$

Graphical analysis suggests the effect of Equation 6.31 on the standard error of estimate to vary as a function of  $\rho[x,y]$ ,  $\rho[x,z]$ ,  $\rho[y,z]$  and  $K$  [i.e. the selection/truncation ratio]. Figures 6.10 - 6.17 depict the reaction of the ratio  $\Delta = \sigma[Y|X]/\sigma[y|x]$  [Delta] to changes in  $\rho[x,y]$ ,  $\rho[x,z]$ ,  $\rho[y,z]$  and  $K$ .

For the purpose of the analysis the predictor and criterion were assumed to be standardised [0; 1] variables. Figures 6.10 - 6.17 seem to suggest that Equation 6.31 generally has the effect of reducing the standard error of estimate. There are, however, conditions where this is clearly not the case. An increase in the standard error of estimate seem to occur at low values of  $\rho[x,z]$  combined with high values on  $\rho[y,z]$ .

An analysis of the effect of corrections for Case 3[i] [Case C[i]] selection on  $Z$ , on selection decision making based on the conditional probability of success, should focus on the relative position of  $Y_c$  in the conditional criterion distribution. Since the critical criterion cutoff value,  $Y_c$ , is defined on a scale unaffected by [systematic and/or random] measurement error, the relative position of  $Y_c$  in the conditional criterion distribution depends on the reaction of both the expected criterion performance and the standard error of estimate to corrections for Case 3[i] [Case C[i]] selection on  $Y$ . A factor complicating the issue, however, is the complexity of the reaction of both the expected criterion performance and the standard error of estimate to corrections for Case 3[i] [Case C[i]] selection on  $Y$  over the space defined by  $\rho[x,y]$ ,  $\rho[y,z]$ ,  $\rho[x,z]$  and  $K$ . Since  $\beta[Y|X]$  can be either bigger or smaller than  $\beta[y|x]$  and  $\sigma[Y|X]$ , similarly, can be either bigger or smaller than  $\sigma[y|x]$ , at least four possible outcomes have to be considered. The probabilities of each of these four possible outcomes occurring are not equal.



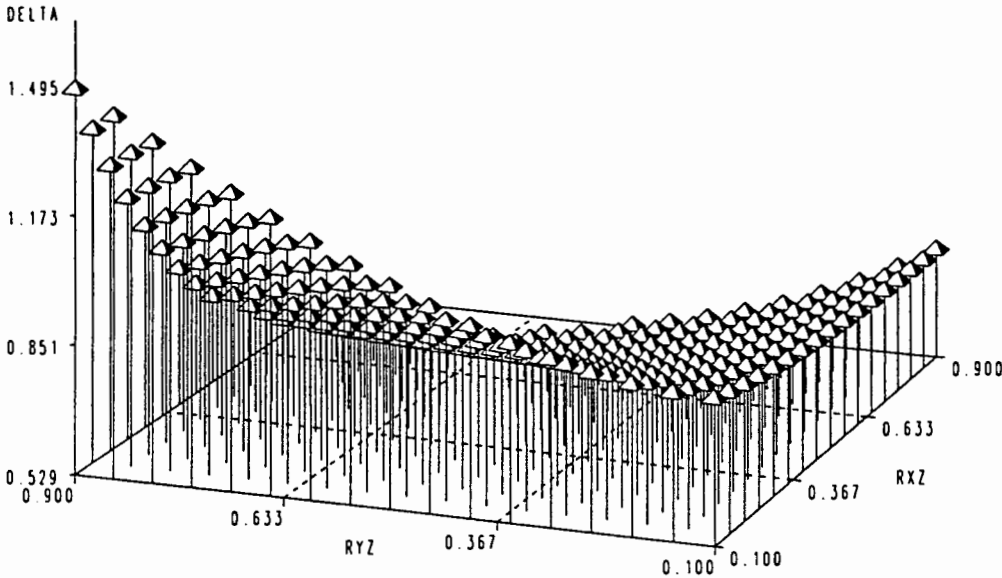


Figure 6.10: The ratio  $\Delta = \sigma[Y|X]/\sigma[y|x]$  [Delta], under Case 3[i] [Case C[i]] restriction of range as a function of  $\rho[x,z]$  and  $\rho[y,z]$  for  $\rho[x,y]$  fixed at 0.10 and K fixed at 2

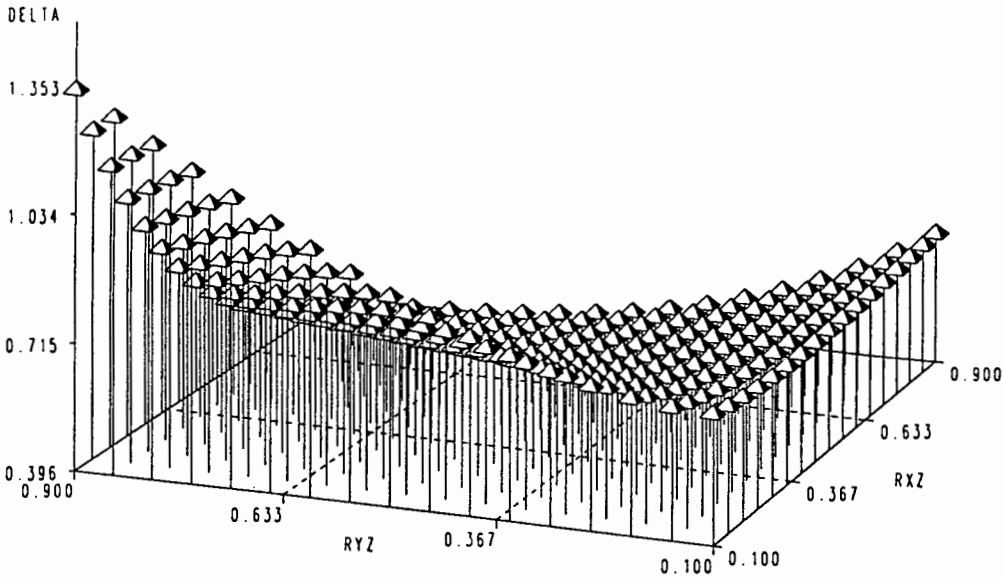


Figure 6.11: The ratio  $\Delta = \sigma[Y|X]/\sigma[y|x]$  [Delta], under Case 3[i] [Case C[i]] restriction of range as a function of  $\rho[x,z]$  and  $\rho[y,z]$  for  $\rho[x,y]$  fixed at 0.30 and K fixed at 2



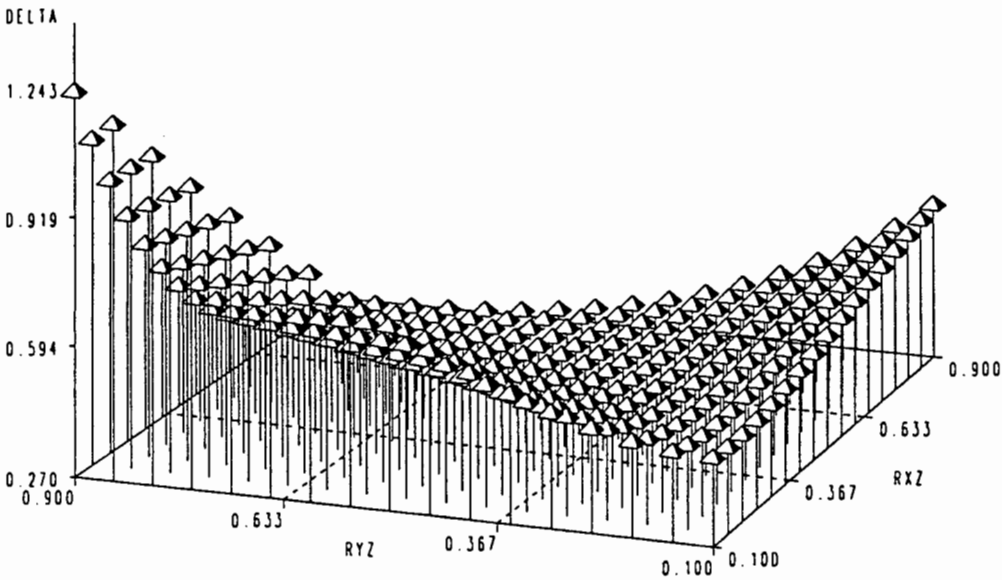


Figure 6.12: The ratio  $\Delta = \sigma[Y|X]/\sigma[y|x]$  [Delta], under Case 3[i] [Case C[i]] restriction of range as a function of  $\rho[x,z]$  and  $\rho[y,z]$  for  $\rho[x,y]$  fixed at 0.60 and K fixed at 2

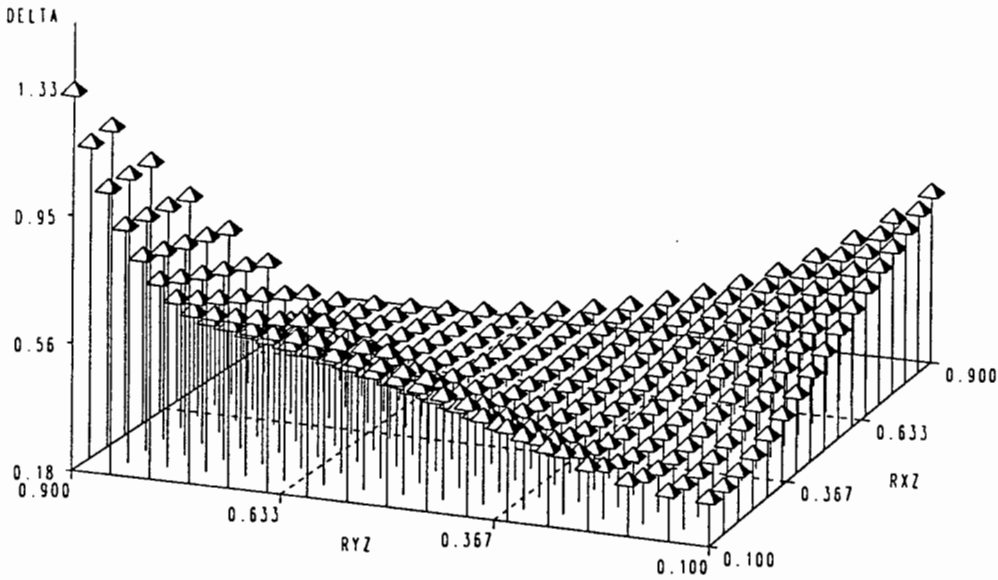


Figure 6.13: The ratio  $\Delta = \sigma[Y|X]/\sigma[y|x]$  [Delta], under Case 3[i] [Case C[i]] restriction of range as a function of  $\rho[x,z]$  and  $\rho[y,z]$  for  $\rho[x,y]$  fixed at 0.80 and K fixed at 2

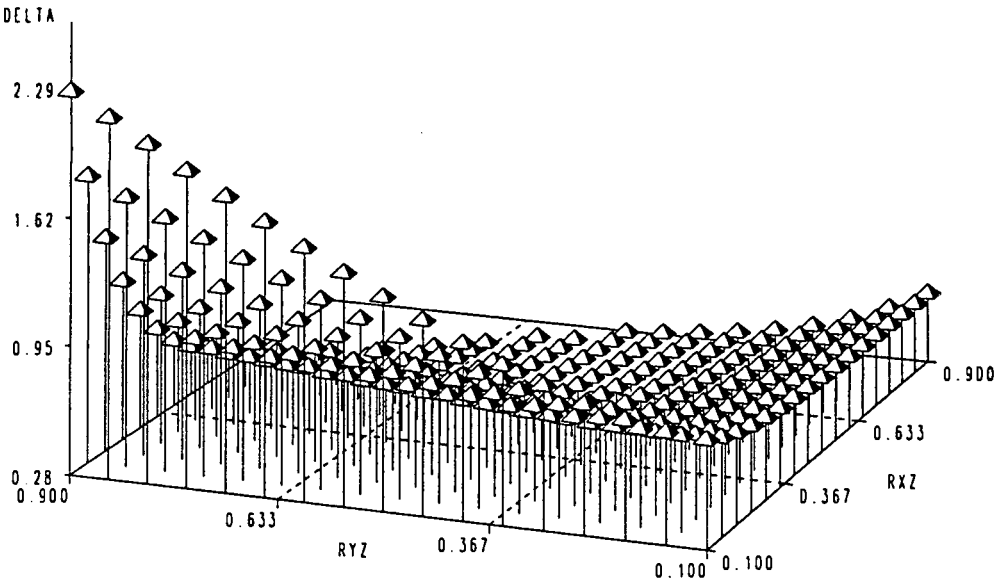


Figure 6.14: The ratio  $\Delta = \sigma[Y|X]/\sigma[y|x]$  [Delta], under Case 3[i] [Case C[i]] restriction of range as a function of  $\rho[x,z]$  and  $\rho[y,z]$  for  $\rho[x,y]$  fixed at 0.10 and K fixed at 4

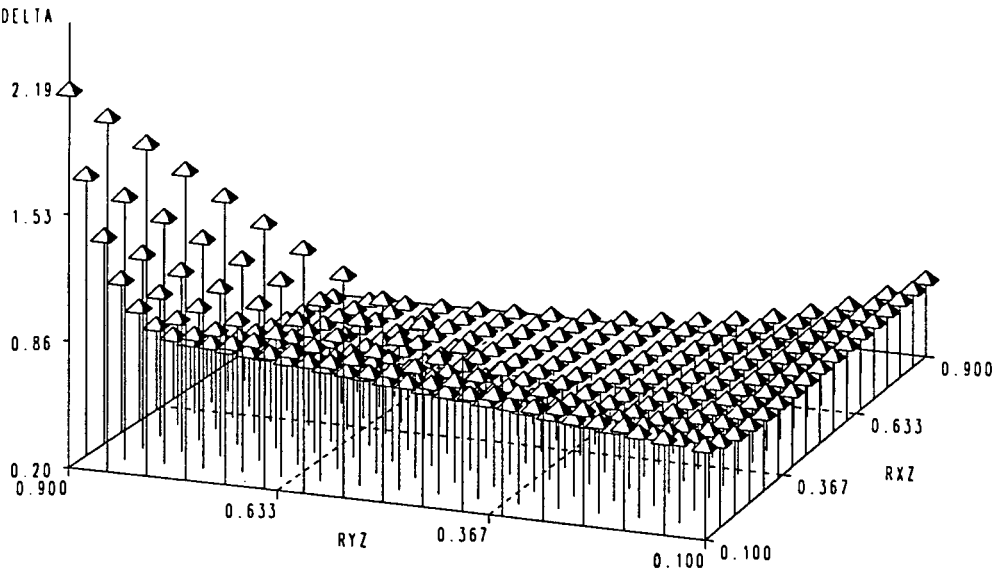


Figure 6.15: The ratio  $\Delta = \sigma[Y|X]/\sigma[y|x]$  [Delta], under Case 3[i] [Case C[i]] restriction of range as a function of  $\rho[x,z]$  and  $\rho[y,z]$  for  $\rho[x,y]$  fixed at 0.30 and K fixed at 4

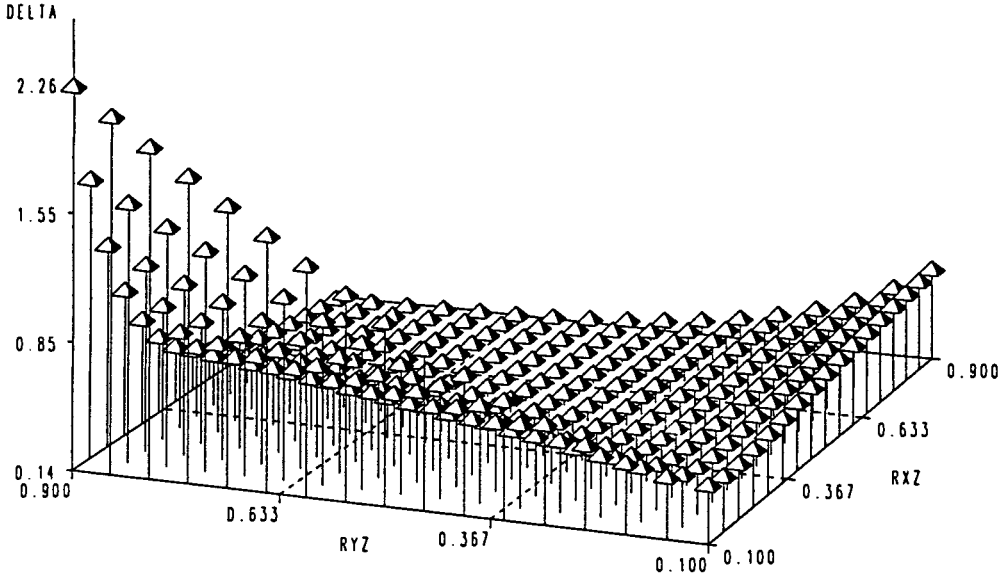


Figure 6.16: The ratio  $\Delta = \sigma[Y|X]/\sigma[y|x]$  [Delta], under Case 3[i] [Case C[i]] restriction of range as a function of  $\rho[x,z]$  and  $\rho[y,z]$  for  $\rho[x,y]$  fixed at 0.60 and K fixed at 4

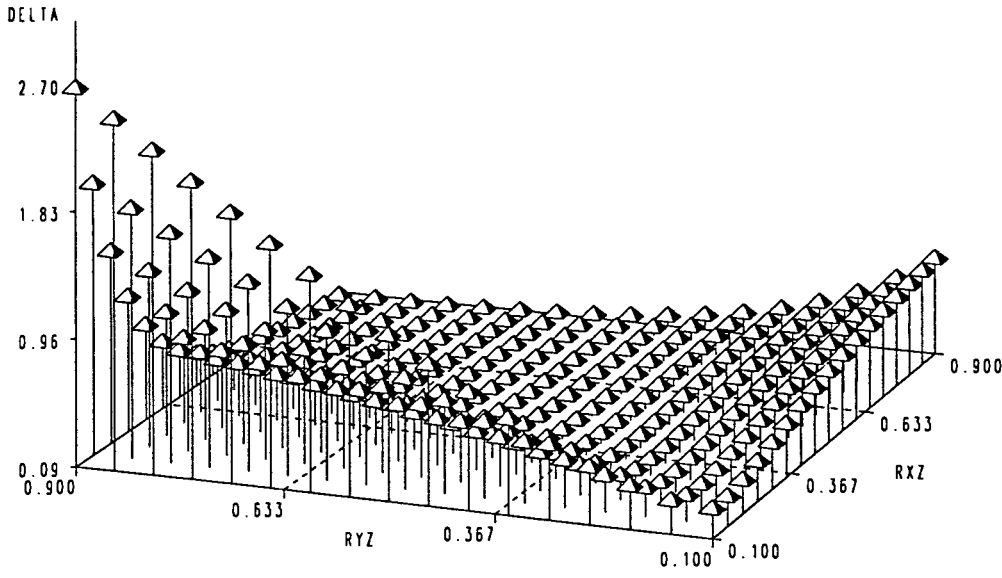


Figure 6.17: The ratio  $\Delta = \sigma[Y|X]/\sigma[y|x]$  [Delta], under Case 3[i] [Case C[i]] restriction of range as a function of  $\rho[x,z]$  and  $\rho[y,z]$  for  $\rho[x,y]$  fixed at 0.80 and K fixed at 4

The most likely outcome seems to be an increase in slope combined with a decrease in the standard error of estimate. A highly unlikely outcome, on the other hand, seems to be a decrease in slope combined with an increase in the standard error of estimate. The position of  $X_i$  relative to the point  $X_s$  where the corrected and uncorrected regression lines intersect [i.e.  $E[Y|X = X_s] = E[y|x = X_s]$ ], furthermore needs to be considered. An analysis of the effect of corrections for Case 3[i] [Case C[i]] selection on Z, on selection decision making based on the conditional probability of success, would therefore have to reckon with the eight possible conditions depicted in Table 6.3.

Table 6.3: The eight possible conditions affecting the effect of corrections for Case 3[i] [Case C[i]] selection on Z on selection decision making based on the conditional probability of success

	$\beta[Y X] < \beta[y x]$		$\beta[Y X] > \beta[y x]$	
	$X_i < X_s$	$X_i > X_s$	$X_i < X_s$	$X_i > X_s$
$\sigma[Y X] < \sigma[y x]$	?	?	P decreases	P increases
$\sigma[Y X] > \sigma[y x]$	P increases	P decreases	?	?

The three dimensions defining Table 6.3 collectively determine whether the change in the expected criterion performance and the change in the variance of the conditional criterion distribution exert a concerted or an antagonistic influence on the conditional probability of success. In the latter case, with the two processes opposing each other, the effect on the conditional probability of success is somewhat more difficult to fathom since it depends on which one of the two processes dominate and thus on  $X_i$  [more specifically, the extent to which  $X_i$  deviates from  $X_s$ ]. Table 6.3 depicts the anticipated effect of Equations 6.29, 6.30 and 6.31 on the conditional probability of success

Correction for Case 3[i] [Case C[i]] selection on Z will [probably] not affect selection decision-making based on conditional probability of success when selection quotas limit the number of applicants required. The initial rank-order based on  $P[y \geq Y_c|x]$  will not be altered by Equations 6.29, 6.30 and 6.31. The same  $N_s$  applicants will therefore still be selected as long as the principle of strict top down selection is still adhered to.

In the case of unrestricted selection, however, decisions will be affected. Since  $Y_c$  is assumed constant, the change in the conditional probability of success attributable to Equations 6.29, 6.30 and 6.31 would probably move a number of applicants across the critical criterion cutoff  $\alpha_s$ . The direction of the migration would depend on the three dimensions of Table 6.3. Either an increase or a decrease in the number of applicants selected could therefore occur

### 6.2.3 Effect Of The Joint Correction For Case 2 [Case A] Restriction Of Range And Criterion Unreliability On The Decision Function

In the typical validity study, restriction of range and random measurement error are functioning in unison to yield an attenuated validity coefficient that could severely underestimate the actual operational validity [Lee, Millar & Graham, 1982; Schmidt, Hunter & Urry, 1976;]. Expressions that provide a joint correction of the validity coefficient for direct restriction of range on the predictor and random measurement error in the criterion [Equation 3.10 and Equation 3.13] had been presented earlier [Bobko, 1983; Mendoza & Mumford, 1987]. For the purpose of the ensuing discussion Case 2 [Case A] selection on X will also be assumed. The following assumption therefore apply [Gulliksen, 1950]:

- $\beta[Y|X] = \beta[y|x]$ ; and
- $\sigma[Y|X] = \sigma[y|x]$

The subsequent discussion will also be restricted to the effect of removal of random measurement error from the criterion only. Equation 3.2 provides an expression for the validity coefficient corrected for the attenuating effect of criterion unreliability only.

Equation 6.32 examines the effect of the joint correction for Case 2 [Case A] selection on X and criterion unreliability on the slope parameter of the regression of  $T_Y$  on X.

By assumption the regression of Y on X is not affected by Case selection on X:

$$\beta[Y|X] = \beta[y|x] \text{-----} 6.32.1$$

Substituting Equation 6.3 in Equation 6.32.1:

$$\rho[X,Y](\sigma[Y]/\sigma[X]) = \rho[x,y](\sigma[y]/\sigma[x]) \text{-----} 6.32.2$$

Isolating the unrestricted criterion standard deviation in Equation 6.32.2:

$$\sigma[Y] = (\rho[x,y]\sigma[y]\sigma[X]) / (\sigma[x]\rho[X,Y]) \text{-----} 6.32.3$$

The slope of the regression of the infallible or true unrestricted criterion scores on the observed and unrestricted predictor scores can be written as:

$$\beta[T_Y|X] = \rho[T_Y,X](\sigma[T_Y]/\sigma[X]) \text{-----} 6.32.4$$

Substituting Equation 3.2 and Equation 2.16 in Equation 6.32.4:

$$\begin{aligned}\beta[T_Y | X] &= (\rho[X, Y] / \sqrt{\rho_{tt}Y}) (\sqrt{\rho_{tt}Y} \sigma[Y] / \sigma[X]) \\ &= (\rho[X, Y]) (\sigma[Y] / \sigma[X]) \text{-----} 3.32.5\end{aligned}$$

Substituting Equation 3.32.3 in Equation 3.32.5:

$$\begin{aligned}\beta[T_Y | X] &= (\rho[X, Y]) \{ (\rho[x, y] \sigma[y] \sigma[X]) / (\sigma[x] \rho[X, Y]) \} (1 / \sigma[X]) \\ &= \rho[x, y] (\sigma[y] / \sigma[x]) \\ &= \beta[y | x] \text{-----} 6.32\end{aligned}$$

Equation 6.32 indicates that the slope parameter is not affected by the joint correction for Case 2 [Case A] selection on X and criterion unreliability.

Equation 6.33 examines the effect of the joint correction for Case 2 [Case A] selection on X and criterion unreliability on the intercept parameter of the regression of  $T_Y$  on X.

By assumption the regression of the unrestricted criterion on the unrestricted predictor is not affected by Case 2 [Case A] selection on X:

$$\alpha[Y | X] = \alpha[y | x] \text{-----} 6.33.1$$

Based on Equation 6.2, the intercept of the regression of the unrestricted criterion true scores on the unrestricted, observed predictor can be expressed as:

$$\alpha[T_Y | X] = \mu[T_Y] - \beta[T_Y | X] \mu[X] \text{-----} 6.33.2$$

However, since,  $\mu[T_Y] = \mu[Y - E] = \mu[Y] - \mu[E] = \mu[Y]$ , Equation 6.33.2 can be rewritten as:

$$\alpha[T_Y | X] = \mu[Y] - \beta[T_Y | X] \mu[X] \text{-----} 6.33.3$$

Combining Equation 6.32 with the assumption that the regression of Y on X is not affected by Case 2 [Case A] selection on X, implies that Equation 6.33.3 can be rewritten as:

$$\begin{aligned}\alpha[T_Y | X] &= \mu[Y] - \beta[Y | X] \mu[X] \\ &= \alpha[y | x] \text{-----} 6.33\end{aligned}$$



The intercept parameter is therefore also not affected by the joint correction for Case 2 [Case A] selection on X and criterion unreliability.

Equations 6.32 and 6.33 imply that the expected criterion performance associated with applicants will not be affected by the simultaneous removal of random measurement error from the criterion and the reversal of explicit Case 2 selection on the predictor. Selection decision-making will consequently not be affected by the simultaneous removal of random measurement error from the criterion and the reversal of explicit Case 2 selection on the predictor, irrespective of whether selection quotas apply or not.

Equation 6.34 examines the effect of the joint correction for Case 2 [Case A] selection on X and criterion unreliability on the standard error of estimate of the regression of  $T_Y$  on X.

Based on Equation 6.5, the unrestricted true score criterion standard deviation conditional on the unrestricted, observed predictor score can be expressed as:

$$\sigma[T_Y|X] = \sigma[T_Y] \sqrt{1 - \rho^2[T_Y, X]} \text{-----} 6.34.1$$

Substituting Equation 2.16 in Equation 6.34.1:

$$\sigma[T_Y|X] = (\sqrt{\rho_{tt} Y_s[Y]}) \sqrt{1 - \rho^2[T_Y, X]} \text{-----} 6.34.2$$

By assumption the regression of Y on X is not affected by Case selection on X:

$$\beta[Y|X] = \beta[y|x] \text{-----} 6.34.3$$

Substituting Equation 6.3 in Equation 6.34.3:

$$\rho[X, Y](\sigma[Y]/\sigma[X]) = \rho[x, y](\sigma[y]/\sigma[x]) \text{-----} 6.34.4$$

Isolating the unrestricted, observed criterion standard deviation:

$$\sigma[Y] = (\rho[x, y]\sigma[y]\sigma[X]) / (\sigma[x]\rho[X, Y]) \text{-----} 6.34.5$$

However, Equation 3.4 provides an expression for the validity coefficient corrected for Case 2 [Case A] selection::

$$\rho[X, Y] = (\sigma[X]\rho[x, y]) / \sqrt{(\sigma^2[X]\rho^2[x, y] + \sigma^2[x] - \sigma^2[x]\rho^2[x, y])} \text{-----} 6.34.6$$

Substituting Equation 6.34.6 in Equation 6.34.4:

$$\begin{aligned}
 \sigma[Y] &= (\rho[x,y]\sigma[y]\sigma[X]) / \{(\sigma[x]\sigma[X]\rho[x,y]) / \sqrt{(\sigma^2[X]\rho^2[x,y] + \sigma^2[x] - \sigma^2[x]\rho^2[x,y])}\} \\
 &= (1/\sigma[x])(\sigma[y])\sqrt{(\sigma^2[X]\rho^2[x,y] + \sigma^2[x] - \sigma^2[x]\rho^2[x,y])} \\
 &= \sigma[y]\sqrt{((\sigma^2[X]/\sigma^2[x])\rho^2[x,y] + 1 - \rho^2[x,y])} \\
 &= \sigma[y]\sqrt{(1 - \rho^2[x,y] + \rho^2[x,y]K^2)} \text{-----} 6.34.7
 \end{aligned}$$

Gulliksen [1950, p. 124] provides the following expression for the criterion reliability coefficient corrected for Case 2 [Case A] selection:

$$\rho_{ttY} = 1 - (\sigma^2[y]/\sigma^2[Y])(1 - \rho_{tty}) \text{-----} 6.34.8$$

Multiplying Equation 6.34.8 with  $\sigma^2[Y]$  and simplifying:

$$\begin{aligned}
 \rho_{ttY}\sigma^2[Y] &= \{1 - (\sigma^2[y]/\sigma^2[Y])(1 - \rho_{tty})\}\sigma^2[Y] \\
 &= \sigma^2[Y] - \sigma^2[y](1 - \rho_{tty}) \\
 &= \sigma^2[Y] - \sigma^2[y] + \sigma^2[y]\rho_{tty} \text{-----} 6.34.9
 \end{aligned}$$

Substituting Equation 6.34.7 in Equation 6.34.9:

$$\begin{aligned}
 \rho_{ttY}\sigma^2[Y] &= \sigma^2[y](1 - \rho^2[x,y] + \rho^2[x,y]K^2) - \sigma^2[y] + \sigma^2[y]\rho_{tty} \\
 &= \sigma^2[y]\{(1 - \rho^2[x,y] + \rho^2[x,y]K^2) - 1 + \rho_{tty}\} \\
 &= \sigma^2[y]\{\rho_{tty} - \rho^2[x,y] + \rho^2[x,y]K^2\} \text{-----} 6.34.10
 \end{aligned}$$

Let  $K = \sigma[X]/\sigma[x]$

Equation 3.13 jointly corrects  $\rho[x,y]$  for Case 2 [Case A] restriction of range and criterion unreliability:

$$\rho^2[T_Y, X] = (\rho^2[x,y]K^2) / (\rho_{tty} - \rho^2[x,y] + \rho^2[x,y]K^2) \text{-----} 6.34.11$$

Substituting Equations 6.34.10 and 6.34.11 in the square of Equation 6.34.2:

$$\begin{aligned}
 \sigma^2[T_Y | X] &= (\rho_{ttY}\sigma[Y])(1 - \rho^2[T_Y, X]) \\
 &= \sigma^2[y]\{\rho_{tty} - \rho^2[x,y] + \rho^2[x,y]K^2\} \{1 - (\rho^2[x,y]K^2) / (\rho_{tty} - \rho^2[x,y] + \rho^2[x,y]K^2)\} \\
 &= \sigma^2[y](\rho_{tty} - \rho^2[x,y] + \rho^2[x,y]K^2) - \sigma^2[y]\rho^2[x,y]K^2 \\
 &= \sigma^2[y](\rho_{tty} - \rho^2[x,y] + \rho^2[x,y]K^2 - \rho^2[x,y]K^2) \\
 &= \sigma^2[y](\rho_{tty} - \rho^2[x,y]) \text{-----} 6.34.12
 \end{aligned}$$

Taking the square root of Equation 6.34.12:

$$\begin{aligned}\sigma[T_Y|X] &= \sigma[y]\sqrt{(\rho_{ttY} - \rho^2[x,y])} \\ &\neq \sigma[y|x] \\ &\neq \sigma[Y|X] \text{-----6.34}\end{aligned}$$

Equation 6.34 returns the same result as Equation 6.11. The effect of the joint correction for Case 2 [Case A] selection on X and criterion unreliability on the criterion variance conditional on X, is therefore equivalent to the effect of the correction for criterion unreliability only. The value Equation 6.34 returns for  $\sigma[T_Y|X]$  therefore decreases from the value for  $\sigma[y|x]$  calculated from fallible criterion data on the selected applicant population.

The critical question is how this affects the conditional probability of success. Since a constant  $Y_c$  is assumed, only the reaction of the standard error of estimate needs to be taken into consideration since the parameters of the regression of Y on X [i.e. the expected criterion performance] remain invariant. The decrease in the standard error brought about by the correction will have the effect of translating the critical criterion cutoff to a more extreme Z-score as shown by Equation 6.12. The precise effect on  $P[T_Y \geq Y_c | X = X_i]$ , moreover, will depend on the position of  $X_i$  relative to  $X_c = [Y_c - a]/b$  [i.e.  $P[Y \geq Y_c | X = X_c] = 0.50$ ]. For all  $X_i < X_c$ ,  $P[T_Y \geq Y_c | X = X_i]$  will decrease relative to  $P[Y \geq Y_c | X = X_i]$  and for all  $X_i > X_c$ ,  $P[T_Y \geq Y_c | X = X_i]$  will increase relative to  $P[Y \geq Y_c | X = X_i]$ . Only for  $X_i = X_c$ , will  $P[Y \geq Y_c | X = X_i]$  remain unaffected.

Jointly correcting the standard error of estimate for the attenuating effect of criterion unreliability and Case 2 [Case A] selection on X will, however, very unlikely change the selection decisions on applicants when selection is restricted by quotas. Although the conditional probabilities are altered by the correction as indicated above, the rank-order of the applicants in terms of their chances of success remain exactly the same. Consequently, the same top  $N_v$  would still be selected. Partially correcting the standard error of estimate for the attenuating effect of criterion unreliability will only affect restricted selection if an insufficient number of applicants initially [i.e. prior to corrections] meet the entry requirement  $\alpha_s$ .

In the case of unrestricted selection, however, jointly correcting the standard error of estimate for criterion unreliability and Case 2 [Case A] selection on X could affect the selection decision on applicants by increasing the number of applicants qualifying for selection [assuming  $\alpha_s > 0.50$ ]. Correcting the standard error of estimate via Equation 6.34 has the effect of pushing any conditional probability greater than 0.50 but less than  $\alpha_s$  [assuming  $\alpha_s > 0.50$ ] towards, and possibly past,  $\alpha_s$ , thus increasing the selection ratio.

### 6.3 CONCLUDING COMMENTS

The foregoing results apply only for the joint correction for criterion unreliability and Case 2 [Case A] selection on X. Analogous joint corrections would be possible, however, such as corrections for:

- both criterion and predictor unreliability [i.e. working with  $E[T_X|X]$  instead of  $X$ ] and Case 2 [Case A] selection on  $X$ ;
- criterion unreliability and Case C selection on  $C$ ; and
- both criterion and predictor unreliability and Case C selection on  $Z$ ;

The effect of these joint corrections on the information on applicants relevant to the strategy matrix has not been examined. It would almost certainly differ from the results reported here.

In chapter 5 the question was raised whether statistical corrections for random measurement error and/or various forms of restriction of range affect statistical significance testing on the corrected correlation/validity coefficient, and if so, in what way. An equivalent question could and should be raised with regards to the parameters of the regression of the criterion on the predictor.

No attempt is made in this study to formally examine this issue. The hypothesis is, however, put forward, that, as in the case of the uncorrected correlation coefficient and the uncorrected slope coefficient, the exceedence probabilities associated with the corrected correlation coefficient and the corrected slope coefficient would coincide.

## 6.4 SUMMARY

Chapter 6 presented the logic underlying selection decision making so as to establish the necessary theoretical foundation for the subsequent analysis. It was then argued that the logic underlying the corrections for restriction of range and/or criterion unreliability applied to the validity coefficient should be extended to the decision rule in terms of which applicants are screened for employment.

The subsequent analyses derived analytical expressions indicating the effects of applying correction for restriction of range and/or criterion unreliability to the slope, intercept and conditional variance parameters of the regression of Y on X. The effect of changes to these parameters on actual selection decision-making was then investigated.

## CHAPTER 7

### CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER RESEARCH

The first objective of the ensuing discussion is to review the reasoning in terms of which the necessity and significance of the foregoing research was justified, to provide a concise description of the nature of the analyses, and to present a summary of the findings of the analyses. Following on from this, the second objective of the discussion is to present recommendations on further research required.

#### 7.1 CONCLUSIONS

##### 7.1.1 Reviewing The Necessity And Significance Of The Research

Selection, as it is traditionally interpreted, represents a critical human resource intervention in any organisation in as far as it regulates the movement of employees into, through and out of the organisation. As such selection thus firstly represents a potentially powerful instrument through which the human resource function can add value to the organisation [Boudreau, 1983b; Boudreau & Berger, 1985a; Cascio, 1991b; Cronshaw & Alexander, 1985]. Selection, however, also represents a relatively visible mechanism through which access to employment opportunities are regulated. Because of this latter aspect, selection, more than any other human resource intervention, has been singled out for intense scrutiny from the perspective of fairness and affirmative action [Arvey & Faley, 1988; Milkovich & Boudreau, 1994;]. Two basic criteria are implied in terms of which selection procedures need to be evaluated, namely efficiency and equity [Milkovich & Boudreau, 1994]. The quest for efficient and equitable selection procedures requires periodic psychometric audits to provide the feedback needed to refine the selection procedure towards greater efficiency and to provide the evidence required to vindicate the organisation should it be challenged in terms of anti-discriminatory legislation. According to the Guidelines for the validation and use of personnel selection procedures [Society for Industrial Psychology, 1992], the Principles for the validation and use of personnel selection procedures [Society for Industrial and Organisational Psychology, 1987] and the Kleiman and Faley [1985] review of selection litigation, such a psychometric audit of a selection procedure would require the human resource function to demonstrate that:

- the selection procedure has its foundation in a scientifically credible performance theory;
- the selection procedure constitutes a business necessity; and



- the manner in which the selection strategy combines applicant information can be considered fair.

The empirical evidence needed to meet the aforementioned burden of persuasion is acquired through a simulation of the actual selection procedure on a sample taken from the applicant population. Internal and external validity constitute two criteria in terms of which the credibility and convincingness of the evidence produced by such a simulation would be evaluated. The following two crucial questions are thereby indicated:

- to what extent can the researcher be confident that the research evidence produced by the selection simulation corroborates the latent structure/nomological network postulated by the research hypothesis within the limits set by the specific conditions characterising the simulation?; and
- to what extent can the researcher be confident that the conclusions reached on the simulation will generalise or transport to the area of actual application?

The conditions under which selection procedures are typically simulated and those prevailing at the eventual use of a selection procedure normally differ to a sufficient extent to challenge the transportability of the validation research evidence. Nevertheless, given the applied nature of selection validation research, an attempt at generalisation is unavoidable. According to Stanley and Campbell [1963] external validity is threatened by the potential specificity of the demonstrated effect of the independent variable[s] to particular features of the research design not shared by the area of application. In selection validation research the effect of the [composite] independent variable on the criterion is captured by the validity coefficient. The area of application is characterised by a sample of actual applicants drawn from the applicant population and measured on a battery of fallible predictors with the aim of "estimating their actual contribution to the organisation [i.e. ultimate criterion scores] and not an indicator of it attenuated by measurement error" [Campbell, 1991, p. 694]. The estimate is derived from a weighted linear composite of predictors derived from a representative sample of the actual applicant population. The question regarding external validity, in the context of selection validation research, essentially represents an inquiry into the unbiasedness of the parametric validity coefficient estimate [i.e. the sample statistic] obtained through the validation study. The parameter of interest is the correlation coefficient obtained when the sample weights derived from a representative sample are applied to the applicant population and the weighted composite score is correlated with the criterion, unattenuated by measurement error, in the population [Campbell, 1991]. The preceding discussion clearly identifies the term "applicant population" to be of central importance should a sufficiently precise depiction of the area of actual application be desired. The term "applicant population", however, even if defined as the population to which a selection procedure will be applied, still has an annoying impreciseness to it. A more unambiguous definition of the term, however, depends on how the selection procedure is positioned relative to any selection requirements already in



use [i.e. whether it would replace, follow on, or be integrated with current selection requirements]. This issue, moreover, is linked to the question regarding the appropriate decision alternative with which to compare the envisaged selection procedure when examining its strategic merit.

In the context of selection validation research, given the aforementioned depiction of the area of application, the following specific threats to external validity can be identified [Campbell, 1991; Lord & Novick, 1968; Tabachnick & Fidell, 1989]:

- the extent to which the actual or operationalised criterion contains random measurement error;
- the extent to which the actual or operationalised criterion is systematically biased; i.e. the extent to which the actual criterion is deficient and/or contaminated [Blum & Naylor, 1968];
- the extent to which the validation sample is an unrepresentative, biased, sample from the applicant population in terms of homogeneity and specific attributes [e.g. motivation, knowledge/experience];
- the extent to which the sample size and the ratio of sample size to number of predictors allow capitalisation on chance and thus overfitting of the data.

The conditions listed as threats all affect the validity coefficient [Campbell, 1991; Crocker & Algina, 1986; Dobson, 1988; Hakstian, Schroeder & Rogers, 1988; Lord & Novick, 1968; Mendoza & Mumford, 1987; Messick, 1989; Olsen & Becker, 1983; Schepers, 1996], some consistently exerting upward pressure, others downward pressure and for some the direction of influence varies. It thus follows that, to the extent that the aforementioned threats operate in the validation study but do not apply to the actual area of application, the obtained validity coefficient cannot, without formal consideration of these threats, be generalised to the actual area of application. Thus, the obtained validity coefficient cannot, without appropriate corrections, be considered an unbiased estimate of the actual validity coefficient of interest.

Statistical corrections to the validity coefficient are generally available to estimate the validity coefficient that would have been achieved had it been calculated under the condition that characterise that area of actual application [Gulliksen, 1950; Pearson, 1903; Thorndike, 1949]. Campbell [1991, p. 701] consequently recommends that:

If the point of central interest is the validity of a specific selection procedure for predicting performance over a relatively long time period for the population of job applicants to follow, then it is necessary to correct for restriction of range, criterion unreliability, and the fitting of error by differential predictor weights. Not to do so is to introduce considerable bias into the estimation process.

The remainder of the argument in terms of which a selection procedure is developed and justified could, however, also be biased by any discrepancy between the conditions under which the selection procedure is simulated and those prevailing during the eventual use of the selection procedure. Relatively little concern, however, seem to exist for the transportability of the decision function derived from the selection simulation and descriptions/assessments of selection decision utility and fairness. This seems to be a somewhat strange state of affairs. The external validity problems of validation designs are reasonably well documented [Barrett, Philips & Alexander, 1981; Cook, Campbell & Peracchio, 1992; Guion & Cranny, 1982; Sussmann & Roberson, 1986]. It is therefore not as if the psychometric literature is unaware of the problem of generalising validation study research findings to the eventual area of application. The decision function is probably the pivot of the selection procedure because it firstly captures the underlying performance theory, but more importantly from a practical perspective, because it guides the actual acceptance and rejection choices of applicants [i.e. it forms the basis of the strategy matrix]. Restricting the statistical corrections to the validity coefficient would leave the decision function unaltered even though it might also be distorted by the same factors affecting the validity coefficient. Basically the same logic also applies to the evaluation of the decision rule in terms of selection utility and fairness. Correcting only the validity coefficient would leave the "bottom-line" evaluation of the selection procedure unaltered. Restricting the statistical corrections to the validity coefficient basically means that practically speaking nothing really changes.

The general objective of the research reported here is therefore to determine whether:

- specific discrepancies between the conditions under which the selection procedure is simulated and those prevailing during the eventual use of the selection procedure produces bias in estimates required to specify and justify the procedure;
- to delineate appropriate statistical corrections of the validity coefficient, decision rule and descriptions/assessments of selection decision utility and fairness, required to align the contexts of evaluation/validation and application; and
- to determine whether the corrections should be applied in validation research.

With reference to this latter aspect the following argument is pursued. The evaluation of any personnel intervention in essence constitutes a process where information is obtained and analysed/processed at a cost with the purpose of making a decision [i.e. choosing between two or more treatments] which results in outcomes with certain value to the decision maker. To add additional information to the evaluation/decision process and/or to extend the analyses of information could be considered rational if it results in an increase in the value of the outcomes at a cost lower than the increase in value. The foregoing argument thus implies that corrections applied to the obtained correlation coefficient are rational to the extent that [Boudreau, 1991]:

- the corrections change decisions on:

- ❖ the validity of the research hypothesis [or at least the a priori probability of rejecting  $H_0$  assuming  $H_0$  to be false]; and/or
  - ❖ the choice of applicants to select; and/or
  - ❖ the appropriate selection strategy option; and/or
  - ❖ the fairness of a particular selection strategy.
- the change in decisions have significant consequences; and
  - the cost of applying the statistical corrections are low.

The argument is thus by implication that there is little merit in applying statistical corrections should they not change any part of the total case built by the validation research team in defense of the selection procedure even if they should rectify systematic bias in the obtained estimates.

### 7.1.2 Research Objectives

The specific research objectives addressed in this study are:

- to determine the individual and combined effect of corrections for attenuation due to the unreliability of the criterion only and restriction of range on the magnitude of the validity coefficient;
- to determine the individual and combined effect of corrections for attenuation due to the unreliability of the criterion only and restriction of range on the magnitude of the standard error of the validity coefficient;
- to determine the individual and combined effect of corrections for attenuation due to the unreliability of the criterion only and restriction of range on the empirically derived exceedence probabilities  $\alpha$ , or achieved significance level];
- to determine the individual and combined effect of corrections for attenuation due to the unreliability of the criterion only and restriction of range on the a priori probabilities  $[1-\beta]$  or power of the tests of the significance for the validity coefficient;
- to determine the individual and combined effect of corrections for attenuation due to the unreliability of the criterion only and restriction of range on the parameters

[intercept, slope and conditional criterion variance] of the linear regression of the criterion on the predictor;

- to determine the individual and combined effect of corrections for attenuation due to the unreliability of the criterion only and restriction of range on the magnitude of the expected criterion performance conditional on the level of predictor performance; and
- to determine the individual and combined effect of corrections for attenuation due to the unreliability of the criterion only and restriction of range on the magnitude of the probability of a substandard criterion performance conditional on the level of predictor performance.

### 7.1.3 Methodology

A combination of analytical [i.e. mathematical/algebraic] and computational approaches were utilised in pursuit of the aforementioned research objectives. The computational analyses consisted of series of solutions of the various correction equations in which the factors affecting a parameter of interest were systematically varied. The behaviour of the parameter of interest was then subsequently graphically mapped over the space defined by the relevant determinants. The SAS software system was used for all analyses.

## 7.1.4 Summary Of Results

### 7.1.4.1 The Effect Of Separate And Combined Statistical Corrections For Restriction Of Range And Random Measurement Error On The Magnitude Of The Validity Coefficient

#### 7.1.4.1.1 The Effect Of The Correction For Criterion Unreliability Only On The Magnitude Of The Validity Coefficient

Equation 3.2 provides an expression for the partially disattenuated validity coefficient. The partially disattenuated validity coefficient  $\rho[X, T_Y]$  is a function of the attenuated validity coefficient  $\rho[X, Y]$  and the criterion reliability coefficient  $\rho_{ttY}$

$$\rho[X, T_Y] = \rho[X, Y] / \sqrt{\rho_{ttY}} \text{-----} 3.2$$

Correcting the Pearson correlation coefficient for unreliability of the criterion measures only has the effect of increasing the magnitude of the validity coefficient. The extent to which  $\rho[X, T_Y]$  exceeds  $\rho[X, Y]$  increases curvilinearly as  $\rho_{ttY}$  decreases over all positive values of  $\rho[X, Y]$ . The partially disattenuated validity coefficient is a linear function of the attenuated validity coefficient with the slope of the function equal to the inverse of the square root of the reliability coefficient. The utility of applying Equation 3.2 at the upper end of the  $\rho_{ttY}$  -scale consequently becomes questionable since the gain in the magnitude of the validity coefficient approaches zero as  $\rho_{ttY}$  approaches unity.

#### 7.1.4.1.2 The Effect Of The Correction For Case 1 [Case B] Restriction Of Range On The Magnitude Of The Validity Coefficient

Equation 3.3 corrects the Pearson correlation coefficient for Case 1 [Case B] selection on X. The corrected validity coefficient  $\rho[X, Y]$  is a function of the restricted validity coefficient  $\rho[x, y]$  and the severity of the restriction of range on the variable on which truncation occurs [i.e. the magnitude of  $K = \sigma[y] / \sigma[Y]$ ].

$$\rho[X, Y] = \sqrt{(1 - K^2(1 - \rho^2[x, y]))} \text{-----} 3.3$$



Correcting the validity coefficient  $\rho[x,y]$  for Case 1 [Case B] restriction of range has the effect of increasing the magnitude of the validity coefficient. The extent to which  $\rho[X,Y]$  exceeds  $\rho[x,y]$  increases as the severity of the restriction of range increases [i.e. as  $K = \sigma[y]/\sigma[Y]$  decreases]. The corrected validity coefficient seems to be practically a linear function of  $\rho[x,y]$  under conditions of extreme restriction of range [i.e., small values of  $K = \sigma[y]/\sigma[Y]$ ]. However, the function gradually seems to transform into a definite positively accelerating curvilinear function as  $K$  approach unity. Under conditions of no restriction of range [i.e.  $K = 1$ ],  $\rho[X,Y]$  necessarily equals  $\rho[x,y]$  and consequently the function again returns to a linear function. The effect of Case 1 [Case B] restriction of range, furthermore, seems to decrease as  $\rho[X,Y]$  increases.

#### 7.1.4.1.3 The Effect Of The Correction For Case 2 [Case A] Restriction Of Range On The Magnitude Of The Validity Coefficient

Equation 3.4 provides an expression for the validity coefficient corrected for Case 2 [Case A] restriction of range. The unrestricted validity coefficient  $\rho[X,Y]$  is a function of the restricted validity coefficient  $\rho[x,y]$  and the truncation ratio [i.e. the magnitude of  $K = \sigma[X]/\sigma[x]$ ].

$$\rho[X,Y] = K\rho[x,y] / \{K^2\rho^2[x,y] + 1 - \rho^2[x,y]\}^{1/2} \text{-----} 3.4$$

Correcting the restricted validity coefficient for Case 2 [Case A] restriction of range increases the magnitude of the correlation. The unrestricted validity coefficient is a curvilinear function of the restricted coefficient. The degree of curvilinearity increases as restriction of range becomes more severe [i.e. as  $K = \sigma[X]/\sigma[x]$  increases from unity]. Correcting extreme high and extreme low  $\rho[x,y]$  values for Case 2 [Case A] selection consequently produces only very small improvements, while restricted correlations in the middle range, especially as  $K$  increases from unity, are more strongly affected. Over the full range of unrestricted validity coefficient values, the effect of Case 2 [Case A] selection on  $X$  increases as  $K$  increases.

#### 7.1.4.1.4 The Effect Of The Correction For Case C Restriction Of Range On The Magnitude Of The Validity Coefficient

Equation 3.5 provides a correction formula to correct for selection bias in the validity coefficient under Case 3[i] [Case C[i]] conditions. The unrestricted validity coefficient  $\rho[X,Y]$  is a function of the restricted validity coefficient  $\rho[x,y]$ , the restricted correlation between the predictor and the variable on



which explicit selection occurs  $\rho[x,z]$ , the restricted correlation between the criterion and the variable on which explicit selection occurs  $\rho[y,z]$  and the truncation ratio [i.e. the magnitude of  $K = \sigma[Z]/\sigma[z]$ ].

$$\rho[X,Y] = \{\rho[x,z]\rho[y,z]K^2\} / [\sqrt{\{1-\rho^2[x,z] + \rho^2[x,z]K^2\}}\sqrt{\{1-\rho^2[y,z] + \rho^2[y,z]K^2\}}] \text{-----} 3.5$$

The effect of Case 3[i] [Case C[i]] selection on Z on the restricted validity coefficient increases as the selection that occurs on Z becomes more severe [i.e. as  $K = \sigma[Z]/\sigma[z]$  increases] and as the correlation between the variable on which explicit selection occurs and the criterion, as well as the selection variable and the predictor, increases. The unrestricted validity coefficient  $\rho[X,Y]$  appears to remain a linear function of  $\rho[x,y]$  over all values of K,  $\rho[x,z]$  and  $\rho[y,z]$ . The relationship between the unrestricted validity coefficient  $\rho[X,Y]$  and K, however, seems to become increasingly curvilinear as  $\rho[x,z]$  and  $\rho[y,z]$  increases.

#### 7.1.4.1.5 The Effect Of The Joint Correction For Case 2 [Case A] Restriction Of Range On X And Criterion Unreliability On The Magnitude Of The Validity Coefficient

Equation 3.13 provides a joint correction of the validity coefficient for Case 2 [Case A] restriction of range directly on the predictor and the unreliability of the predictor. The corrected validity coefficient  $\rho[X,T_Y]$  is a function of the restricted and attenuated validity coefficient  $\rho[x,y]$ , the criterion reliability coefficient and the truncation ratio on the predictor [i.e.  $K = \sigma[X]/\sigma[x]$ ].

$$\rho[X,T_Y] = \sqrt{\{(\rho^2[x,y])K^2\}} / \{K^2\rho^2[x,y] - \rho^2[x,y] + \rho_{tty}\} \text{-----} 3.13$$

Applying Equation 3.13 to  $\rho[x,y]$  increases the magnitude of the validity coefficient over the complete space defined by K,  $\rho[x,y]$  and  $\rho_{tty}$ . The extent to which Case 2 [Case A] selection on X and criterion unreliability jointly affect  $\rho[X,T_Y]$  increases as  $K = \sigma[X]/\sigma[x]$  and  $\rho_{tty}$  increases. The relationship between  $\rho[X,T_Y]$  and  $\rho_{tty}$  becomes increasingly curvilinear at the lower end of the  $\rho[x,y]$ -scale, as K increases. The relationship between  $\rho[X,T_Y]$  and  $\rho[x,y]$  changes from linear, when  $K = 1$ , into an increasingly curvilinear relationship as K increases.

#### 7.1.4.2 The Effect Of Separate And Combined Statistical Corrections For Restriction Of Range And Random Measurement Error On The Magnitude Of The Standard Error Of The Pearson Correlation Coefficient

##### 7.1.4.2.1 The Effect Of The Correction For Criterion Unreliability Only On The Magnitude Of The Standard Error Of The Pearson Correlation Coefficient

Equation 4.22 provides the expression for the square of the standard error of the partially disattenuated validity coefficient derived by Kelly [1947, p. 529]  $\sigma^2[\rho^*[X, T_Y]]$  is dependent on the attenuated validity coefficient  $\rho[X, Y]$ , the split-half reliability coefficient  $\rho'_{ttY}$  and the number of cases included in the validation group.

$$\sigma^2[\rho^*[X, T_Y]] = \rho^2[X, T_Y]/(n-2)(\rho^2[X, T_Y] + \rho^{-1}[X, Y] - \rho'^{-1}_{ttY} + 4\rho'^{-2}_{ttY} - 5/4) \text{-----} 4.22$$

The standard error of the partially disattenuated validity coefficient decreases curvilinearly as the reliability of the criterion increases, increases sharply as the attenuated correlation increases at the lower end of the criterion reliability scale but decreases slowly as the attenuated correlation increases at the upper end of the criterion reliability scale. The standard error of the partially disattenuated validity coefficient furthermore decreases as the sample size increases.

The standard error of the partially disattenuated validity coefficient is greater than the standard error of the attenuated validity coefficient in the whole space defined by  $\rho[X, Y]$  and  $\rho_{ttY}$ , except for a small region characterised by high  $\rho[X, Y]$  and  $\rho_{ttY}$  values. This region in which the standard error of the partially disattenuated validity coefficient decreases relative to the standard error of the attenuated validity coefficient gradually seems to expand to lower  $\rho[X, Y]$  values as  $n$  increases. This finding might, however, be attributable to the known weakness of the expression [Equation 4.13] used to represent the standard error of the attenuated validity coefficient in the analysis towards the upper end of the  $\rho[X, Y]$  scale.

Equations 4.23 – 4.25 provide three alternative expressions of the standard error of the partially disattenuated validity coefficient derived by Bobko and Rieck [1980, p. 388].

$$\sigma^2[\rho^*[X, T_Y]] = [A + B - C] \text{-----} 4.23$$

$$\sigma^2[\rho^*[X, T_Y]] = [A + B] \text{-----} 4.24$$

$$\sigma^2[\rho^{\wedge}[X,T_Y]] = [A] \text{-----} 4.25$$

Where:

$$A = n^{-1}(\rho_{tt}Y)^{-1}(1-\rho^2[X,Y])^2;$$

$$B = (0,25n)^{-1}(\rho_{tt}Y)^{-3}(\rho[X,Y])^2(1-\rho_{tt}Y)^2; \text{ and}$$

$$C = n^{-1}(\rho_{tt}Y)^{-2}\rho[X,Y]\{\rho_{tt}Y(1-2\rho^2[X,Y])-(0,5\rho^2[X,Y])(1-2\rho^2[X,Y]-\rho^2_{tt}Y)\}$$

The conditions under which each expression becomes appropriate were described earlier.

The standard error of the partially disattenuated validity coefficient calculated through Equations 4.23 and 4.24 tends to increase curvilinearly as the attenuated correlation increases at low values of  $\rho_{tt}Y$  but this trend gradually reverses as  $\rho_{tt}Y$  increases, until, at the upper end of the criterion reliability scale, the standard error moderately decreases as  $\rho[X,Y]$  increases. The standard error produced by Equations 4.23 and 4.24, furthermore, increase as the criterion reliability decreases, peaks at the maximum  $\rho[X,Y]$  value possible, given  $\rho_{tt}Y$  and decreases as sample size increases. The first two expressions derived by Bobko and Rieck [1980] therefore tend to behave essentially in the same way as the expression derived by Kelly [1947].

The standard error of the partially disattenuated validity coefficient calculated through Equation 4.25 also tends to increase as  $\rho_{tt}Y$  decreases and also decreases as  $n$  increases. However, the standard error obtained through Equation 4.25 reacts differently to an increase in the attenuated correlation. Although still curvilinearly related to  $\rho[X,Y]$ , the standard error now seems to gradually decrease as  $\rho[X,Y]$  increases for all values of  $\rho_{tt}Y$ .

For Equations 4.23, 4.24 and 4.25, the standard error of the partially disattenuated validity coefficient exceeds the standard error of the attenuated validity coefficient for all permissible combinations of values of  $\rho[X,Y]$ ,  $\rho_{tt}Y$  and  $n$ . For the first two standard error expressions derived by Bobko and Rieck [1980] the reaction of ratio of the standard errors of the partially disattenuated and attenuated validity coefficients [ $V = \sigma[\rho^{\wedge}_c]/\sigma[\rho^{\wedge}]$ ] to changes in  $\rho[X,Y]$ ,  $\rho_{tt}Y$  and  $n$  is very similar. The ratio  $V$  increases as  $\rho[X,Y]$  increases and  $\rho_{tt}Y$  decreases. The ratio  $V$  peaks both at high  $\rho_{tt}Y$  and high  $\rho[X,Y]$  and at low  $\rho_{tt}Y$  combined with low-moderate  $\rho[X,Y]$ .

For Equation 4.25, however,  $V$  remains constant over changes in  $\rho[X,Y]$  for all values of  $\rho_{tt}Y$  and increases as  $\rho_{tt}Y$  decreases. This would imply that the standard error of the attenuated validity coefficient shadows/tails the standard error of the partially disattenuated validity coefficient over all

values of  $\rho[X,Y]$  and  $\rho_{tt}Y$ , but that the latter tend to gain on the former as the criterion reliability decreases.

#### 7.1.4.2.2 The Effect Of The Correction For Case 1 [Case B] Restriction Of Range On The Magnitude Of The Standard Error Of The Pearson Correlation Coefficient

No formal derivation of a specific expression for the standard error of the correlation coefficient corrected for Case 1 [Case B] selection could be uncovered. The reaction of the standard error of the correlation coefficient to Case 1 [Case B] selection was therefore not examined.

#### 7.1.4.2.3 The Effect Of The Correction For Case 2 [Case A] Restriction Of Range On The Magnitude Of The Standard Error Of The Pearson Correlation Coefficient

Equation 4.35 provides the expression for the square of the standard error of the validity coefficient corrected for Case 2 [Case A] restriction of range on X derived by Bobko and Rieck [1980, p. 390]:

$$\sigma^2[\rho^*[X,Y]] = \{K^2(1-\rho^2[x,y]+\rho^2[x,y]K^2)^{-3}\} \{(1-\rho^2[x,y])^2/n\} \text{-----} 4.35$$

Where:  $K = (\sigma[X]/\sigma[x])$

The standard error of the validity coefficient corrected for Case 2 [Case A] restriction of range seems to increase as the restricted validity coefficient decreases and to increase relatively sharply as K increases at low values of  $\rho[x,y]$ , but to decrease slowly as K increases at high values of  $\rho[x,y]$ . As would be expected, the standard error increases as n decreases.

The ratio of the standard error of the validity coefficient corrected for Case 2 [Case A] restriction of range to the standard error of the observed, restricted validity coefficient [i.e.  $V = \sigma[\rho^*]/\sigma[\rho^*]$ ] seems to copy/imitate the behaviour of the standard error of the corrected coefficient as described above. The standard error of the validity coefficient corrected for Case 2 [Case A] restriction of range generally is greater than the standard error of the observed, uncorrected validity coefficient. However, V does decrease below unity in a relatively small region of the  $\rho[x,y], K$  plane characterised by high  $\rho[x,y]$  and K values. When estimating  $\sigma[\rho^*]$  through Equation 4.13,  $\sigma[\rho^*]$  estimated through Equation 4.35 will thus exceed the former at small selection ratios and high restricted [i.e. uncorrected] correlations.

#### 7.1.4.2.4 The Effect Of The Correction For Case C Restriction Of Range On The Magnitude Of The Standard Error Of The Pearson Correlation Coefficient

Equation 4.36 provides the expression for the standard error of the validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range derived by Allen and Dunbar [1990, p. 80].

$$\sqrt{\sigma^2[\rho^{\wedge}[X,Y]]} = \{ \rho'^2[X,Y]\sigma^2[\rho^{\wedge}[x,y]] + \rho'^2[X,Z]\sigma^2[\rho^{\wedge}[x,z]] + \rho'^2[Y,Z]\sigma^2[\rho^{\wedge}[y,z]] + 2\rho'[X,Y]\rho'[X,Z]\sigma[\rho^{\wedge}[x,y],\rho^{\wedge}[x,z]] + 2\rho'[X,Y]\rho'[Y,Z]\sigma[\rho^{\wedge}[x,y],\rho^{\wedge}[y,z]] + 2\rho'[X,Z]\rho'[Y,Z]\sigma[\rho^{\wedge}[x,z],\rho^{\wedge}[y,z]] \}^{1/2} \quad 4.36$$

Where:

$\rho'[X,Y]$  indicates the partial derivative of the adjusted correlation  $\rho[X,Y]$  with respect to the unadjusted correlation  $\rho[x,y]$ ; specifically

$$\rho'[X,Y] = (1 + W\rho^2[x,z])^{-1/2} (1 + W\rho^2[y,z])^{-1/2}$$

$$\rho'[X,Z] = W(1 + W\rho^2[x,z])^{-3/2} (1 + W\rho^2[y,z])^{-1/2} (\rho[y,z] - \rho[x,y]\rho[x,z])$$

$$\rho'[Y,Z] = W(1 + W\rho^2[x,z])^{-1/2} (1 + W\rho^2[y,z])^{-3/2} (\rho[x,z] - \rho[x,y]\rho[y,z]); \text{ and}$$

$$W = (\sigma^2[Z]/\sigma^2[z]) - 1$$

The two large sample estimates of the sampling variance of a correlation corrected for indirect [Case 3[i] [Case C[i]]] selection proposed by Allen and Dunbar [1990] differ in terms of the equations chosen for the estimation of the asymptotic variance and covariance components. The difference fundamentally lies in the assumptions on which these estimates are based. Allen and Dunbar [1990, p. 86] distinguish between two sets of assumptions, namely:

- set A; "error/n to second and third powers is negligible, linearity of regression, and mesokurtosis of the joint distribution of explicit and implicit variables; and
- set B; error/n to second and third powers is negligible and linearity of regression."

Only the variance and covariance equations based on the more stringent assumption set A are presented below.

$$\sigma[\rho^{\wedge}[x,y],\rho^{\wedge}[x,z]] = (1/n) \{ \rho[y,z](1 - \rho^2[x,z] - \rho^2[x,y])^{-1/2} \rho[x,y]\rho[x,z](1 - \rho^2[y,z] - \rho^2[x,y] - \rho^2[x,z]) \}$$

$$\sigma[\rho^{\wedge}[x,y],\rho^{\wedge}[y,z]] = (1/n)\{\rho[x,z](1-\rho^2[x,y]-\rho^2[y,z])^{-1/2}\rho[x,y]\rho[y,z](1-\rho^2[x,z]-\rho^2[x,y]-\rho^2[y,z])\}$$

$$\sigma[\rho^{\wedge}[x,z],\rho^{\wedge}[y,z]] = (1/n)\{\rho[x,y](1-\rho^2[x,z]-\rho^2[y,z])^{-1/2}\rho[x,z]\rho[y,z](1-\rho^2[x,y]-\rho^2[x,z]-\rho^2[y,z])\}$$

$$\sigma^2[\rho^{\wedge}[x,z]] = (1 - \rho^2[x,z])^2/n$$

$$\sigma^2[\rho^{\wedge}[y,z]] = (1 - \rho^2[y,z])^2/n$$

$$\sigma^2[\rho^{\wedge}[x,y]] = (1 - \rho^2[x,y])^2/n$$

The standard error of the validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range tends to increase as K increases. The standard error tends to increase as  $\rho[x,y]$  decreases, but with the rate of increase dependent on  $\rho[x,z]$ ,  $\rho[y,z]$  and K. The rate of increase initially accelerates as  $\rho[x,z]$  increases and then decreases again as  $\rho[x,z]$  increases further. The point of maximum slope, relative to the  $\rho[x,y]$  axis, on the  $\rho[x,z]$  axis shifts upward on the latter axis as  $\rho[y,z]$  increases. The rate of increase tends to decelerate as K increases.

The standard error of the validity coefficient corrected for Case 3[i] [Case C[i]] restriction of range generally tends to be larger than the standard error of the uncorrected validity coefficient. The standard error of the uncorrected validity coefficient can, however, under certain conditions exceed the standard error of the corrected coefficient. The conditions favoring  $V=\sigma[\rho^{\wedge}_c]/\sigma[\rho^{\wedge}] < 1$  seem to be extreme truncation [i.e. high K values], higher correlations between the selection variable Z and the criterion and predictor variables respectively [in the selected group], and lower  $\rho[x,y]$  values. The region of maximum difference in the magnitude of the standard errors seems to occur at high levels of  $\rho[x,y]$  and a reciprocal combination of high  $\rho[x,z]$  and low  $\rho[y,z]$  or vice versa.

#### 7.1.4.2.5 The Effect Of The Joint Correction For Case 2 [Case A] Restriction Of Range On X And Criterion Unreliability On The Magnitude Of The Standard Error Of The Pearson Correlation Coefficient

Equation 4.37 provides an expression for the square of the standard error of the validity coefficient jointly corrected for Case 2 [Case A] restriction of range and criterion unreliability, derived by Bobko [1983].



$$\sigma[\hat{\rho}[X, T_Y]] = (n)^{-1}(\rho_{ttY})^{-1}(K^2 D^3) \{ (1 - \rho^2[x, y])^2 + (1/4)\rho^2[x, y]\rho^{-2}_{ttY}(1 - \rho^2_{ttY})^2 - \{\rho[x, y]\rho^{-1}_{ttY}\} \{ \rho[x, y](1 - \rho^2[x, y] - \rho^2_{ttY}) - (1/2)\rho[x, y]\rho_{ttY}(1 - 2\rho^2[X, Y] - \rho^2_{ttY}) \} \} \quad 4.37$$

where,  $D = (1 - \rho^2[X, Y]\rho^{-1}_{ttY} + K^2\rho^2[X, Y]\rho^{-1}_{ttY})$

and

$$K = \sigma[X]/\sigma[x]$$

The standard error of the validity coefficient jointly corrected for Case 2 [Case A] restriction of range and criterion unreliability tends to increase curvilinearly as  $\rho_{ttY}$  decreases and increases as  $K$  increases. The standard error of the corrected validity coefficient, furthermore decreases as sample size increases and tends to increase as  $\rho[x, y]$  decreases. The standard error of the corrected validity coefficient peaks at low values of  $\rho_{ttY}$  and  $\rho[x, y]$ .

The reaction of the standard error of the corrected validity coefficient to changes in the relevant parameters seem to closely correspond to the behaviour of the standard error of the uncorrected coefficient over the same parameter space. The standard error of the validity coefficient jointly corrected for Case 2 [Case A] restriction of range and criterion unreliability generally tends to be larger than the standard error of the uncorrected validity coefficient. The standard error of the uncorrected validity coefficient can, however, under certain conditions exceed the standard error of the corrected coefficient. The conditions favoring  $V = \sigma[\hat{\rho}_c]/\sigma[\hat{\rho}] < 1$  seem to be restricted to a small region of the  $\rho_{ttY}, \rho[x, y]$  plane characterised by high values on both the  $\rho_{ttY}$  and  $\rho[x, y]$  axis.

#### **7.1.4.3 The Effect Of Separate And Combined Statistical Corrections For Restriction Of Range And Random Measurement Error On The Magnitude Of The Empirically Derived Exceedence Probability $\alpha_B$**

##### **7.1.4.3.1 The Effect Of The Correction For Criterion Unreliability Only On The Magnitude Of The Empirically Derived Exceedence Probability**

The Bobko and Rieck [1980] expressions for the standard error of the partially disattenuated validity coefficient were used to assess the impact of the correction for criterion unreliability on the empirically derived exceedence probability.

##### **7.1.4.3.1.1 The Reliability Coefficient Given A Priori By Theoretical Assumption Or Previously Accepted Knowledge**

The empirically derived exceedence probability  $\alpha_B$  is not affected by the correction for criterion unreliability. The conditional probability of observing the corrected validity coefficient under  $H_0$  is the same as the conditional probability of observing the uncorrected validity coefficient under  $H_0$ . The statistical significance tests for the uncorrected and corrected validity tests are asymptotically identical if  $\rho_{ttY}$  is known a priori.

##### **7.1.4.3.1.2 The Reliability Coefficient Obtained From An Independent Data Set**

If the reliability coefficient used to correct the validity coefficient for the unreliability of the criterion is obtained from an independent data set, and the appropriate expression for the standard error is subsequently used, the empirical exceedence probability is affected. The probability of observing the corrected correlation in a sample of size  $n$  drawn from a population where the two variables being correlated are, in fact, independent is consistently larger than the conditional probability of observing the attenuated coefficient in a sample of the same size under  $H_0$ .

The increase in  $\alpha_B$  produced by the partial correction for attenuation has the effect of changing some significant correlations into insignificant partially disattenuated correlations. The change in decision

occurs in a specific region of the  $\rho[X,Y]$  axis as a function of sample size. The region shifts lower down on the  $\rho[X,Y]$  axis as sample size increases.

#### **7.1.4.3.1.3 The Reliability Coefficient Obtained From The Same Data Set**

If the reliability coefficient used to correct the validity coefficient for the unreliability of the criterion is obtained from the same data set, and the appropriate expression for the standard error is subsequently used, the empirical exceedence probability is affected. . The probability of observing the corrected correlation in a sample of size  $n$  drawn from a population where the two variables being correlated are, in fact, independent is consistently larger than the conditional probability of observing the attenuated coefficient in a sample of the same size under  $H_0$ . The nature of the effect is very similar to those found for the Bobko and Rieck [1980] Case II expression as described in the previous paragraph.

The increase in  $\alpha_b$  produced by the partial correction for attenuation has the effect of changing some significant correlations into insignificant partially disattenuated correlations. The change in decision occurs in a specific region of the  $\rho[X,Y]$  axis as a function of sample size. The region shifts lower down on the  $\rho[X,Y]$  axis as sample size increases.

#### **7.1.4.3.2 The Effect Of The Correction For Case 1 [Case B] Restriction Of Range On The Magnitude Of The Empirically Derived Exceedence Probability**

No formal derivation of a specific expression for the standard error of the correlation coefficient corrected for Case 1 [Case B] selection could be uncovered. The effect of the correction for Case 1 [Case B] restriction of range on the magnitude of the empirically derived exceedence probability could, therefore, not be examined.

#### **7.1.4.3.3 The Effect Of The Correction For Case 2 [Case A] Restriction Of Range On The Magnitude Of The Empirically Derived Exceedence Probability**

The conditional probability of observing the corrected validity coefficient under  $H_0$  equals the conditional probability of observing the uncorrected validity coefficient under  $H_0$  when  $K = \sigma[X]/\sigma[x] = 1$ . This is a logically necessary result, since  $\rho^{\wedge}[x,y] = \rho^{\wedge}[X,Y]$  and  $\sigma[\rho^{\wedge}[x,y]] = \sigma[\rho^{\wedge}[X,Y]]$  when  $K =$

1. For all possible combinations of  $n$ ,  $\rho^{[x,y]}$  and  $K > 1$ , the probability of observing the corrected validity coefficient under  $H_0$  is consistently smaller than the probability of observing the corresponding uncorrected validity coefficient value under  $H_0$ . The effect of the correction for Case 2 [Case A] restriction of range on  $\alpha_b$  intensifies curvilinearly as the uncorrected sample correlation  $\rho^{[x,y]}$  and  $K$  increases.

The decrease in the conditional probability  $\alpha_b$  has the effect of changing some insignificant uncorrected correlations into significant corrected correlations. This effect seems to be relatively sensitive to sample size. At  $n \geq 120$  Case 2 [Case A] restriction of range corrections to the validity coefficient no longer has the effect of changing the significance test outcome [given a significance level of 0.05 and a one-tailed test]. The change in decision occur in a specific, restricted region in the  $\rho^{[x,y]}$ ,  $K$  space as a function of sample size  $n$ . as sample size increases, the region shifts to lower values of  $\rho^{[x,y]}$ , but simultaneously also higher values of  $K$ .

#### 7.1.4.3.4 The Effect Of The Correction For Case 3[i] [Case C[i]] Restriction Of Range On The Magnitude Of The Empirically Derived Exceedence Probability

Correcting the validity coefficient for Case 3[i] [Case C[i]] restriction of range produces either a decrease or an increase in the conditional probability of observing the corrected validity coefficient under  $H_0$ :  $\rho[X,Y] = 0$  relative to the conditional probability of observing the uncorrected validity coefficient under  $H_0$ :  $\rho[x,y] = 0$ . These changes in  $\alpha_b$  brought about by the correction for restriction of range has the effect of changing some insignificant uncorrected correlations into significant corrected correlations and vice versa.

The region in which correlations change from insignificant to significant seem to occur at the lower end of the  $\rho^{[x,y]}$  axis for moderate to high  $\rho^{[x,z]}$  values. This region seems to shift towards zero on the  $\rho^{[x,y]}$  axis as  $K$  increases. The region in which the decision on significance of the validity coefficient is affected, shrinks as  $n$  increases.

The region in which correlations change from significant to insignificant seem to occur at the lower end of the  $\rho^{[x,y]}$  axis for moderate to high  $\rho^{[x,y]}$  values. This region seems to shift towards the upper end of the  $\rho^{[x,y]}$  axis as  $\rho^{[y,z]}$  increases and to spread towards zero on the  $\rho^{[x,z]}$  axis as  $K$  decreases. The region in which the decision on significance of the validity coefficient is affected, shrinks as  $n$  increases.

The conditional probability of observing the corrected validity coefficient under  $H_0: \rho[X, Y] = 0$  seems to decrease relative to the conditional probability of observing the uncorrected validity coefficient under  $H_0: \rho[x, y] = 0$  as  $K$  increases, as the correlation of the predictor with the selection variable  $Z$  increases, as the correlation of the criterion with the predictor increases and as  $\rho^*[x, y]$  decreases.

#### 7.1.4.3.5 The Effect Of The Joint Correction For Case 2 [Case A] Restriction Of Range On X And Criterion Unreliability On The Magnitude Of The Empirically Derived Exceedence Probability

For all combinations of  $n$ ,  $\rho^*[x, y]$ ,  $\rho_{\pi Y}$  and  $K > 1$ , the probability of observing the uncorrected validity coefficient in a sample of size  $n$ , conditional on  $H_0: \rho[x, y] = 0$ , is consistently bigger than the probability of observing the corresponding double corrected validity coefficient, conditional on the hypothesis that  $\rho[X, T_Y] = 0$ . This effect of the joint correction for Case 2 [Case A] restriction of range and criterion unreliability on  $\alpha_b$  intensifies as  $\rho^*[x, y]$  and  $K$  increases and  $\rho_{\pi Y}$  decreases.

The decrease in  $\alpha_b$ , produced by the joint correction, has the effect of changing some insignificant uncorrected correlations into significant corrected correlations when  $K > 1$ . This effect seems to be fairly sensitive to sample size. At  $n \geq 120$ , double corrections to the validity coefficient no longer has the effect of changing the hypothesis test outcome.

The change in the hypothesis test outcome seems to occur in the lower region of the  $\rho^*[x, y]$  axis and the full range of  $\rho_{\pi Y}$  values as a function of  $K$  and sample size  $n$ . As sample size increases, the region shifts to lower values of  $\rho^*[x, y]$  and also seems to become more restricted to the lower region of the  $\rho_{\pi Y}$  axis.

#### **7.1.4.4 The Effect Of Separate And Combined Statistical Corrections For Restriction Of Range And Random Measurement Error On The Magnitude Of The A Priori Probability Of Rejecting $H_0$ If $H_0$ Is False**

##### **7.1.4.4.1 The Effect Of The Correction For Criterion Unreliability Only On The Magnitude Of The A Priori Probability Of Rejecting $H_0$ If $H_0$ Is False**

###### **7.1.4.4.1.1 The Reliability Coefficient Given A Priori By Theoretical Assumption Or Previously Accepted Knowledge**

The power of the statistical significance test of the validity coefficient will not be affected by the correction for criterion unreliability if the standard error of the corrected validity coefficient is calculated by the Bobko and Rieck [1980] Case III expression.

###### **7.1.4.4.1.2 The Reliability Coefficient Obtained From An Independent Sample**

The power of the statistical significance test of the corrected validity coefficient is negatively affected [i.e. the power decreases] if the standard error of the corrected validity coefficient is calculated by the Bobko and Rieck [1980] Case II expression.

###### **7.1.4.4.1.3 The Reliability Coefficient Obtained From The Same Sample**

The power of the statistical significance test of the corrected validity coefficient is negatively affected [i.e. the power decreases] if the standard error of the corrected validity coefficient is calculated by the Bobko and Rieck [1980] Case I expression. The negative effect of Case I partial corrections for attenuation on statistical power increases curvilinearly with increases in the effect size assumed for the uncorrected validity coefficient over all levels of criterion reliability.



#### **7.1.4.4.2 The Effect Of The Correction For Case 1 [Case B] Restriction Of Range On The Magnitude Of The A Priori Probability Of Rejecting $H_0$ If $H_0$ Is False**

No formal derivation of a specific expression for the standard error of the correlation coefficient corrected for Case 1 [Case B] selection could be uncovered. The effect of the correction for Case 1 [Case B] restriction of range on the magnitude of the a priori probability of rejecting the null hypothesis conditional on the null hypothesis being false, could therefore not be examined.

#### **7.1.4.4.3 The Effect Of The Correction For Case 2 [Case A] Restriction Of Range On The Magnitude Of The A Priori Probability Of Rejecting $H_0$ If $H_0$ Is False**

The statistical power of the statistical significance test of the validity coefficient increases if the correction for Case 2 [Case A] restriction of range is applied to the observed, restricted correlation coefficient and the uncorrected effect size estimate.

#### **7.1.4.4.4 The Effect Of The Correction For Case 3[i] [Case C[i]] Restriction Of Range On The Magnitude Of The A Priori Probability Of Rejecting $H_0$ If $H_0$ Is False**

Case 3[i] [Case C[i]] corrections for restriction of range effect statistical power both positively and negatively, depending on the applicable parameter settings. Case 3[i] [Case C[i]] seem to have maximum adverse impact on statistical power when  $\rho^{[x,y]}$  is high and  $\rho^{[x,z]}$  and  $\rho^{[y,z]}$  is low.  $K$  seem to affect power by aggravating the effect of the aforementioned parameters as  $K$  increases. Corrections for Case 3[i] [Case C[i]] restriction of range seem to have maximum beneficial impact on statistical power under those conditions where Case C restrictions have their greatest impact, namely when  $\rho^{[x,y]}$  is low and  $\rho^{[x,z]}$  and  $\rho^{[y,z]}$  is high.  $K$  again seems to affect power by enhancing the effect of the aforementioned parameters as  $K$  increases.

#### 7.1.4.4.5 The Effect Of The Joint Correction For Case 2 [Case A] Restriction Of Range On X And Criterion Unreliability On The Magnitude Of The A Priori Probability Of Rejecting $H_0$ If $H_0$ Is False

The statistical power of the significance test of the validity coefficient is improved by the double correction for restriction of range and criterion unreliability, provided the uncorrected effect size estimate is also corrected via the appropriate formula to obtain the corrected effect size estimate. Maximum power benefits are obtained by the double correction when the initial uncorrected effect size estimate is high,  $\rho_{\alpha Y}$  is low and  $K$  is high. Although  $n$  affects power, it does not seem to affect the change in power brought about by the double correction to the correlation coefficient.

#### 7.1.4.5 The Effect Of Separate And Combined Statistical Corrections For Restriction Of Range And Random Measurement Error On The Magnitude Of The Intercept And Slope Parameters Of The Regression Of The Criterion On The Predictor

##### 7.1.4.5.1 The Effect Of The Correction For Criterion Unreliability Only On The Magnitude Of The Intercept And Slope Parameters Of The Regression Of The Criterion On The Predictor

The regression of the criterion on the predictor remains unaffected by the partial correction for the attenuating effect of the unreliability of the criterion. Neither the Y-axis intercept, nor the X-axis slope are affected by the removal of random measurement error from the criterion.

The expected criterion performance conditional on  $X$  will therefore remain unaffected. The decision on applicant will therefore not change due to the partial correction for attenuation, irrespective of whether selection is restricted by quotas or not.

#### 7.1.4.5.2 The Effect Of The Correction For Case 1 [Case B] Restriction Of Range On The Magnitude Of The Intercept And Slope Parameters Of The Regression Of The Criterion On The Predictor

If selection occurs directly on the predictor X, but the restricted and unrestricted variances are known only for the criterion Y, then by assumption neither the regression of Y on X nor the criterion variance conditional on X will be affected. Thus, by assumption Equations 6.22 and 6.23 apply. The same notation convention as before is used.

$$\begin{aligned}\beta[Y|X] &= \rho[X,Y](\sigma[Y]/\sigma[X]) \\ &= \rho[x,y](\sigma[y]/\sigma[x]) \\ &= \beta[y|x] \text{-----} 6.22\end{aligned}$$

$$\begin{aligned}\sigma^2[Y|X] &= \sigma^2[Y](1 - \rho^2[X,Y]) \\ &= \sigma^2[y](1 - \rho^2[x,y]) \\ &= \sigma^2[y|x] \text{-----} 6.23\end{aligned}$$

Since Equation 6.22 reflects the assumption that  $\mu[Y|X_i]$  is not altered by explicit selection on X, Equation 6.24 also applies.

$$\begin{aligned}\alpha[Y|X] &= \mu[Y] - \beta[Y|X]\mu[X] \\ &= \mu[y] - \beta[y|x]\mu[x] \\ &= \alpha[y|x] \text{-----} 6.24\end{aligned}$$

Although the magnitude of the correlation is influenced by Case 1 [Case B] selection on X, and although the standard errors of the regression coefficients will increase due to this type of Case 1 [Case B] selection, the relevant decision function [and therefore also the contents of the strategy matrix] will not be affected by Case 1 [Case B] selection on X. Corrections to the parameters of the decision function to compensate for this type of Case 1 [Case B] restriction of range are consequently not required.

If, however, the selection occurs directly on the criterion, but the restricted and unrestricted variances are known only for the predictor, then the regression of Y on X is affected. Equations 6.25 and 6.26 demonstrate that corrections of the parameters of the regression of Y on X for the effect of Case 1 [Case B] selection on Y produce an increase in the slope parameter and a decrease in the intercept.

$$\beta[Y|X] = \beta[y|x]\{(1/\rho^2[x,y]) - (K^2/\rho^2[x,y])(1-\rho[x,y])\} \text{-----} 6.25$$

$$\alpha[Y|X] = \mu[Y] - \beta[Y|X]\mu[X] \\ = (\rho[x,y]\sigma[x]/\sigma[y])^{-1}(\mu[X] - \mu[x]) + \mu[y] - \beta[Y|X]\mu[X] \text{-----} 6.26$$

The expression for the regression of Y on X, corrected for Case 1 [Case B] selection on Y, is shown as Equation 6.27.

$$E[Y|X] = (\rho[x,y]\sigma[x]/\sigma[y])^{-1}(\mu[X] - \mu[x]) + \mu[y] + \{\sigma[y]/(\rho[x,y]\sigma[x]) - \\ (\sigma[x]\sigma[y])/(\rho[x,y]\sigma^2[X])(1 - \rho^2[x,y])\}(X_i - \mu[X]) \text{-----} 6.27$$

The change in the regression parameters implies that the regression equation derived from the selected applicant group and the regression equation that would have been obtained from the unrestricted applicant group will intersect at some point  $X_s$  on the X-axis. Consequently, for all  $X_i < X_s$   $E[y|x_i]$  will overestimate  $E[Y|X_i]$  while for all  $X_i > X_s$   $E[y|x_i]$  will underestimate  $E[Y|X_i]$ . Correcting the parameters of the decision function for Case 1 [Case B] selection on Y, consequently will have the effect of elevating the expected criterion performance for all  $X_i > X_s$ , while depressing it for all  $X_i < X_s$ .

The correction of the decision function for Case 1 [Case B] selection on Y will not affect the applicants selected if selection is restricted by quotas and strict top down selection applies. Although  $E[Y|X_i] \neq E[y|x_i]$ , if  $E[y|x_i] > E[y|x_j]$  then  $E[Y|X_i] > E[Y|X_j]$ . The same  $N_v$  applicants would therefore still be selected.

In the case of selection unrestricted by quotas, however, the use of the corrected decision function will affect the selection decision-making. The nature of the effect will depend on the position of  $X_c$  [derived from  $Y_c$  via the decision function calculated on the selected applicant group] relative to  $X_s$ . If  $X_s < X_c$  the number of applicants satisfying the entry requirement [i.e.  $E[Y|X_i] > Y_c$ ] will increase. If  $X_s > X_c$  [unlikely since it would imply a very low  $Y_c$  on a positive criterion], the number of applicants qualifying for selection will decrease.

#### 7.1.4.5.3 The Effect Of The Correction For Case 2 [Case A] Restriction Of Range On The Magnitude Of The Intercept And Slope Parameters Of The Regression Of The Criterion On The Predictor

If Case 2 [Case A] selection occurs directly on the predictor X, then by assumption, neither the regression of Y on X nor the criterion variance conditional on X will be affected. No corrections to the parameters of the regression equation or the standard error of estimate are therefore required. The

regression of X on Y would be affected, but since it is of no real interest in selection validation research, no justification seems to exist to explore it further.

If Case 2 [Case A] selection occurs directly on the criterion Y, then by assumption, neither the regression of X on Y nor the predictor variance conditional on Y will be affected. The regression of Y on X and the criterion variance conditional on X would be affected. The probability of Case 2 [Case A] selection on Y actually occurring in the context of selection validation research, however, seems for all practical purposes to be zero. Although appropriate correction formula therefore should exist, there once more seems to be no practical justification to try and uncover their formulation.

#### 7.1.4.5.4 The Effect Of The Correction For Case 3[i] [Case C[i]] Restriction Of Range On The Magnitude Of The Intercept And Slope Parameters Of The Regression Of The Criterion On The Predictor

Equation 6.29 provides an expression for the slope of the regression of Y on X for the unselected population. Equation 6.29 assumes Case 3[i] [Case C[i]] selection on Z.

$$\beta[Y|X] = \{(\rho[x,y] - \rho[x,z]\rho[y,z]) + (\rho[x,z]\rho[y,z]K^2)\} / \{(1 - \rho[x,z] + \rho^2[x,z]K^2)\} (\sigma[y]/\sigma[x]) \text{-----} 6.29$$

Equation 6.30 provides an expression for the intercept of the regression of Y on X for the unselected population derived from  $\alpha[y|x]$  and Case 3[i] [Case C[i]] selection.

$$\alpha[Y|X] = \{\mu[y] + \rho[y,z](\sigma[y]/\sigma[z])(\mu[Z] - \mu[z])\} - \beta[Y|X]\{\mu[x] + \rho[x,z](\sigma[x]/\sigma[z])(\mu[Z] - \mu[z])\} \text{-----} 6.30$$

Equations 6.29 and 6.30 generally have the effect of increasing the slope of the regression of Y on X relative to the slope of the regression of y on x and, concomitantly, decreasing the intercept. There are, however, parameter settings where the opposite effect is achieved.

The regression of y on x and the estimated/corrected regression of Y on X will intersect at some point  $X_s$  on the X-axis. The magnitude and direction of the change in the expected criterion performance associated with applicants attributable to the application of Equations 6.29 and 6.30 will depend on the position of  $X_i$  relative to  $X_s$  and the magnitude of  $\beta[Y|X]$  relative to  $\beta[y|x]$ . TABLE 6.1 presents a summary of the effect of Equations 6.29 and 6.30 on the expected criterion performance associated with applicants.



Correction for Case 3[i] [Case C[i]] selection on Z will [probably] not affect selection decision-making based on expected criterion performance when selection quotas limit the number of applicants required. The same  $N_s$  applicants will thus still be selected as long as the principle of strict top down selection is adhered to.

In the case of unrestricted selection, however, corrections for Case C selection on Z probably would affect selection decision-making. Since  $Y_c$  is assumed constant, the change in the expected criterion performance attributable to Equations 6.29 and 6.30 would probably move a number of applicants across the critical criterion cutoff. The direction of the migration would once more depend on the position of  $X_i$  relative to  $X_s$  and the magnitude of  $\beta[Y|X]$  relative to  $\beta[y|x]$ . TABLE 6.2 presents a summary of the anticipated effect of Equations 6.29 and 6.30 on the number of applicants fulfilling the entry requirement for selection [i.e.  $E[Y|X] \geq Y_c$ ].

#### 7.1.4.5.5 The Effect Of The Joint Correction For Case 2 [Case A] Restriction Of Range On X And Criterion Unreliability On The Magnitude Of The Intercept And Slope Parameters Of The Regression Of The Criterion On The Predictor

Equation 6.32 depicts the effect of the joint correction for Case 2 [Case A] selection on X and criterion unreliability on the slope parameter of the regression of  $T_Y$  on X.

$$\begin{aligned}\beta[T_Y|X] &= (\rho[X,Y]) \{ (\rho[x,y]\sigma[y]\sigma[X]) / (\sigma[x]\rho[X,Y]) \} (1/\sigma[X]) \\ &= \rho[x,y](\sigma[y]/\sigma[x]) \\ &= \beta[y|x] \text{-----} 6.32\end{aligned}$$

Equation 6.32 indicates that the slope parameter is not affected by the joint correction for Case 2 [Case A] selection on X and criterion unreliability.

Equation 6.33 depicts the effect of the joint correction for Case 2 [Case A] selection on X and criterion unreliability on the intercept parameter of the regression of  $T_Y$  on X.

$$\begin{aligned}\alpha[T_Y|X] &= \mu[Y] - \beta[Y|X]\mu[X] \\ &= \alpha[y|x] \text{-----} 6.33\end{aligned}$$

The intercept parameter is therefore also not affected by the joint correction for Case 2 [Case A] selection on X and criterion unreliability.



Equations 6.32 and 6.33 imply that the expected criterion performance associated with applicants will not be affected by the simultaneous removal of random measurement error from the criterion and the reversal of explicit selection on the predictor. Selection decision-making will consequently not be affected by the simultaneous removal of random measurement error from the criterion and the reversal of explicit selection on the predictor, irrespective of whether selection quotas apply or not.

#### 7.1.4.6 The Effect Of Separate And Combined Statistical Corrections For Restriction Of Range And Random Measurement Error On The Magnitude Of The Standard Error Of Estimate Of The Regression Of The Criterion On The Predictor

##### 7.1.4.6.1 The Effect Of The Correction For Criterion Unreliability Only On The Magnitude Of The Standard Error Of Estimate Of The Regression Of The Criterion On The Predictor

Equation 6.11 provides an expression for the square of the standard error of estimate, corrected for criterion unreliability:

$$\sigma^2[T_Y|X] = \sigma^2[Y][\rho_{ttY} - \rho^2[X,Y]] \leq \sigma^2[Y|X] \text{-----} 6.11$$

The effect of statistically removing random measurement error from the criterion is to reduce the standard error of estimate.

The reduction in the standard error of estimate will, assuming a constant critical criterion cutoff  $Y_c$ , have the effect of moving  $Y_c$  to a more extreme position in the conditional criterion distribution. Consequently, for all  $X_i > X_c = [Y_c - \alpha]/\beta$ ,  $P[T_Y \geq Y_c | X = X_i]$  will increase relative to  $P[Y > Y_c | x = X_i]$ . Conversely, for all  $X_i < X_c = [Y_c - \alpha]/\beta$ ,  $P[T_Y \geq Y_c | X = X_i]$  will decrease relative to  $P[Y > Y_c | x = X_i]$ . For  $X_i = X_c$  the conditional probability will remain unaffected.

Partially correcting the standard error of estimate for the attenuating effect of criterion unreliability will, however, very unlikely change the selection decisions on applicants when selection is restricted by quotas. Although the conditional probabilities are altered by the correction, the rank-order of the applicants in terms of their chances of success remain exactly the same. Consequently, the same top  $N_v$  would still be selected. Partially correcting the standard error of estimate for the attenuating effect

of criterion unreliability will only affect restricted selection if an insufficient number of applicants initially [i.e. prior to corrections] meet the entry requirement  $\alpha_s$ .

In the case of unrestricted selection, however, partially correcting the standard error for criterion unreliability could affect the selection decision on applicants by increasing the number of applicants qualifying for selection [assuming  $\alpha_s > 0.50$ ]. Correcting the standard error for criterion unreliability has the effect of pushing any conditional probability greater than 0.50 but less than  $\alpha_s$  [assuming  $\alpha_s > 0.50$ ] towards, and possibly past,  $\alpha_s$ , thus increasing the selection ratio. Correcting the standard error for criterion unreliability has the effect of pushing the critical predictor cutoff corresponding to the critical probability  $\alpha_s$  towards the point  $X_c$  [as defined above] in the predictor distribution. The selection ratio is consequently increased for any  $\alpha_s > 0.50$  and decreased for any  $\alpha_s < 0.50$ .

#### 7.1.4.6.2 The Effect Of The Correction For Case 1 [Case B] Restriction Of Range On The Magnitude Of The Standard Error Of Estimate Of The Regression Of The Criterion On The Predictor

Case 1 [Case B] selection on X does not affect the standard error of estimate  $\sigma[Y|X]$ . Case 1 [Case B] selection on Y, however, does affect the standard error of estimate  $\sigma[Y|X]$ . Equation 6.28 provides a correction to the standard error of estimate  $\sigma[y|x]$  for Case 1 [Case B] selection on Y.

$$\sigma^2[Y|X] = \sigma[y]\{ \{ (1-\rho^2[x,y]) \} \{ 1 - (K^2(1-\rho^2[x,y])^2)/\rho^2[x,y] \} \} \text{-----} 6.28 \\ \geq \sigma^2[y|x]$$

Correcting  $\sigma[y|x]$  for Case 1 [Case B] selection on Y via Equation 6.28 returns a corrected value  $\sigma[Y|X]$  greater than the obtained standard error of estimate. The difference between the corrected and uncorrected standard errors of estimate seems to be negatively related to the observed correlation  $\rho[x,y]$ .

Correction for Case 1 [Case B] restriction of range on Y effect both the expected criterion performance conditional on X and the standard error of estimate. The effect of the change in expected criterion performance on the conditional probability tends to oppose the effect of the change in the standard error of estimate. The change in the expected criterion performance exerts an upward pressure on the conditional probability for all  $X_i > X_s$  and a downward pressure for all  $X_i < X_s$ . The change in the standard error of estimate, in contrast, produces the opposite effect. The reaction of the conditional probability of success to the aforementioned corrections for Case 1 [Case B] selection on Y,

consequently, depends on the which one of the two processes dominate. However, no single answer to this question exists.

Two points on the X-axis [ $X_{hi}$  &  $X_{lo}$ ] are postulated, positioned in such a way on both sides of  $X_s$ , that  $Z[Y_c; E[y|x=X_{hi}]] = Z[Y_c; E[Y|X=X_{hi}]]$  and  $Z[Y_c; E[y|x=X_{lo}]] = Z[Y_c; E[Y|X=X_{lo}]]$ . The conditional probability of success associated with applicants falling on  $X_{hi}$  or  $X_{lo}$  would therefore not be affected by the preceding corrections for Case 1 [Case B] selection on Y. However, for all  $X_i > X_{hi}$ , and all  $X_i < X_{lo}$ , the effect of the change in expected criterion performance would dominate. Consequently the conditional probability of success associated with those applicants falling above  $X_{hi}$  will increase [assuming  $Y_c > \mu[Y]$ ], while the conditional probability of success associated with those applicants falling below  $X_{lo}$  will decrease. Furthermore, for those applicants located between these two cutoff points [i.e.  $X_{lo} < X_i < X_{hi}$ ], the effect of the change in the standard error of estimate should dominate and consequently the conditional probability should increase.

The application of Equations 6.27 and 6.28 will [probably] not affect selection decision-making based on the conditional probability of success, if such selection is restricted by selection quotas.

The application of Equations 6.27 and 6.28 will, however, affect unrestricted selection decision-making based on the conditional probability of success. The selection ratio should increase since the critical acceptance probability  $\alpha_s$  is constant while the conditional probability of success associated with those applicants falling above  $X_{hi}$  increases. The number of applicants qualifying for selection therefore increases.

#### **7.1.4.6.3 The Effect Of The Correction For Case 2 [Case A] Restriction Of Range On The Magnitude Of The Standard Error Of Estimate Of The Regression Of The Criterion On The Predictor**

The effect of the correction for Case 2 [Case A] restriction of range on the magnitude of the standard error of estimate of the regression of Y on X was not examined.

#### 7.1.4.6.4 The Effect Of The Correction For Case 3[i] [Case C[i]] Restriction Of Range On The Magnitude Of The Standard Error Of Estimate Of The Regression Of The Criterion On The Predictor

The effect of corrections for Case 3[i] [Case C[i]] selection on Z on the standard error of estimate, is depicted by Equation 6.31.

$$\sigma[Y|X] = (\sigma[y]\sqrt{(1 - \rho^2[y,z] + \rho^2[y,z]K^2)}) - \{(\rho[x,y] - \rho[x,z]\rho[y,z]) + (\rho[x,z]\rho[y,z]K^2)\} / \sqrt{(1 - \rho^2[x,z] + \rho^2[x,z]K^2)} \quad \text{-----} \quad 6.31$$

Equation 6.31 generally has the effect of reducing the standard error of estimate. There are, however, conditions where the correction has the opposite effect. An increase in the standard error of estimate seem to occur at low values of  $\rho[x,z]$  combined with high values on  $\rho[y,z]$ .

The effect of corrections for Case 3[i] [Case C[i]] selection on Z, on selection decision making based on the conditional probability of success is fairly complex.

Since the critical criterion cutoff value,  $Y_c$ , is defined on a scale unaffected by [systematic and/or random] measurement error, the relative position of  $Y_c$  in the conditional criterion distribution depends on the reaction of both the expected criterion performance and the standard error of estimate to corrections for Case 3[i] [Case C[i]] selection on Y. A factor complicating the issue, however, is the complexity of the reaction of both the expected criterion performance and the standard error of estimate to corrections for Case 3[i] [Case C[i]] selection on Y over the space defined by  $\rho[x,y]$ ,  $\rho[y,z]$ ,  $\rho[x,z]$  and K. Since  $\beta[Y|X]$  can be either bigger or smaller than  $\beta[y|x]$  and  $\sigma[Y|X]$ , similarly, can be either bigger or smaller than  $\sigma[y|x]$ , at least four possible outcomes have to be considered. The probability of each of these four possible outcomes occurring are not equal. The most likely outcome seems to be an increase in slope combined with a decrease in the standard error of estimate. A highly unlikely outcome, on the other hand, seems to be a decrease in slope combined with an increase in the standard error of estimate. The position of  $X_i$  relative to the point  $X_s$  where the corrected and uncorrected regression lines intersect [i.e.  $E[Y|X = X_s] = E[y|x = X_s]$ ], furthermore needs to be considered. An analysis of the effect of corrections for Case 3[i] [Case C[i]] selection on Z, on selection decision making based on the conditional probability of success, would therefore have to reckon with eight possible conditions.

The three aforementioned aspects collectively determine whether the change in the expected criterion performance and the change in the variance of the conditional criterion distribution exert a concerted or an antagonistic influence on the conditional probability of success. In the latter case, with the two



processes opposing each other, the effect on the conditional probability of success is somewhat more difficult to fathom since it depends on which one of the two processes dominate and thus on  $X_i$  [more specifically, the extent to which  $X_i$  deviates from  $X_s$ ]. Table 6.3 depicts the anticipated effect of Equations 6.29, 6.30 and 6.31 on the conditional probability of success.

Correction for Case 3[i] [Case C[i]] selection on Z will [probably] not affect selection decision-making based on conditional probability of success when selection quotas limit the number of applicants required. The initial rank-order based on  $P[y \geq Y_c | x]$  will not be altered by Equations 6.29, 6.30 and 6.31. The same  $N_s$  applicants will therefore still be selected as long as the principle of strict top down selection is still adhered to.

In the case of unrestricted selection, however, decisions will be affected. Since  $Y_c$  is assumed constant, the change in the conditional probability of success attributable to Equations 6.29, 6.30 and 6.31 would probably move a number of applicants across the critical criterion cutoff  $\alpha_s$ . The direction of the migration would depend on the aforementioned three dimensions. Either an increase or a decrease in the number of applicants selected could therefore occur

#### 7.1.4.6.5 The Effect Of The Joint Correction For Case 2 [Case A] Restriction Of Range On X And Criterion Unreliability On The Magnitude Of The Standard Error Of Estimate Of The Regression Of The Criterion On The Predictor

Equation 6.34 depicts the effect of the joint correction for Case 2 [Case A] selection on X and criterion unreliability on the standard error of estimate of the regression of  $T_Y$  on X.

$$\begin{aligned}\sigma[T_Y | X] &= \sigma[y] \sqrt{(\rho_{T_Y} - \rho^2[x, y])} \\ &\neq \sigma[y | x] \\ &\neq \sigma[Y | X] \text{-----} 6.34\end{aligned}$$

Equation 6.34 returns the same result as Equation 6.11. The effect of the joint correction for Case 2 [Case A] selection on X and criterion unreliability on the criterion variance conditional on X, is therefore equivalent to the effect of the correction for criterion unreliability only. The value Equation 6.38 returns for  $\sigma[T_Y | X]$  therefore decreases from the value for  $\sigma[y | x]$  calculated from fallible criterion data on the selected applicant population.

The decrease in the standard error brought about by the correction, will have the effect of translating the critical criterion cutoff to a more extreme Z-score. The precise effect on  $P[T_Y \geq Y_c | X = X_i]$ ,

moreover, will depend on the position of  $X_i$  relative to  $X_c = [Y_c - a]/b$  [i.e.  $P[Y \geq Y_c | X = X_c] = 0.50$ ]. For all  $X_i < X_c$ ,  $P[Y \geq Y_c | X = X_i]$  will decrease relative to  $P[Y \geq Y_c | X = X_i]$  and for all  $X_i > X_c$ ,  $P[Y \geq Y_c | X = X_i]$  will increase relative to  $P[Y \geq Y_c | X = X_i]$ . Only for  $X_i = X_c$ , will  $P[Y \geq Y_c | X = X_i]$  remain unaffected.

Jointly correcting the standard error of estimate for the attenuating effect of criterion unreliability and Case 2 [Case A] selection on  $X$  will, however, very unlikely change the selection decisions on applicants when selection is restricted by quotas. Although the conditional probabilities are altered by the correction as indicated above, the rank-order of the applicants in terms of their chances of success remain exactly the same. Consequently, the same top  $N_v$  would still be selected. Partially correcting the standard error of estimate for the attenuating effect of criterion unreliability will only affect restricted selection if an insufficient number of applicants initially [i.e. prior to corrections] meet the entry requirement  $\alpha_s$ .

In the case of unrestricted selection, however, jointly correcting the standard error of estimate for criterion unreliability and Case 2 [Case A] selection on  $X$  could affect the selection decision on applicants by increasing the number of applicants qualifying for selection [assuming  $\alpha_s > 0.50$ ]. Correcting the standard error of estimate via Equation 6.34 has the effect of pushing any conditional probability greater than 0.50 but less than  $\alpha_s$  [assuming  $\alpha_s > 0.50$ ] towards, and possibly past,  $\alpha_s$ , thus increasing the selection ratio.

### 7.1.5 Synopsis Of Findings

No unqualified answer exists to the question whether corrections for various forms of restriction of range and/or criterion unreliability should be applied to the validity coefficient, the standard error of the validity coefficient or the parameters of the regression of the criterion on the predictor. Under specific conditions the corrections do affect decisions on the validity of performance hypotheses due to its effect on decisions on the significance of the uncorrected versus the corrected validity coefficient. Under specific conditions the corrections do affect decisions on applicants, especially when selection decisions are not restricted by selection quotas, due to its effect on the slope and intercept parameters of the regression of  $Y$  on  $X$ , and/or due to its effect on the standard error of estimate.



## 7.2 RECOMMENDATIONS FOR FURTHER RESEARCH

Statistical corrections for criterion unreliability and/or restriction of range have been shown to affect the magnitude of the validity coefficient, the probability of observing the validity coefficient under  $H_0$ , the parameters of the regression of the criterion on the predictor [and thus the expected criterion performance conditional on the predictor] and the variance of the conditional criterion distribution [and thus the conditional probability of successful criterion performance].

Two further important facets of the psychometric audit still need to be examined if a comprehensive understanding of the effects of statistical corrections for the various forms of restriction of range and/or criterion unreliability on the defense of a selection procedure is to be achieved. These two facets are selection utility and selection fairness.

### 7.2.1 The Effect Of Statistical Corrections For Restriction Of Range And/or Criterion Unreliability On Utility Assessment

Selection utility models serve as reality simplifying conceptual frameworks designed as aids for "describing, predicting and explaining the usefulness or desirability" of selection decision strategies, "and analysing how that information can be used in decision making" [Boudreau, 1989, p. 228] aimed at improving selection strategies.

Since statistical corrections for criterion unreliability and/or restriction of range have been shown to affect the magnitude of the validity coefficient, the probability of observing the validity coefficient under  $H_0$ , the parameters of the regression of the criterion on the predictor and the variance of the conditional criterion distribution, subsequent research will have to examine the consequence of the effects described previously for the various descriptive indicators of the usefulness or desirability [i.e. utility] of selection procedures. The fundamental objective should therefore firstly be to determine whether it would in any way affect the various utility indicators if these indicators would have been obtained on the unrestricted applicant population [or a representative sample from the applicant population], utilising criterion data free from random measurement error, rather than on a systematically selected sub-population of applicants, utilising criterion data contaminated by random measurement error. If it transpires that the aforementioned differences between the simulated and actual application of a selection strategy do in fact affect the measures/descriptions obtained from the utility indicators, the possibility that appropriate corrections could change the final decision on the selection option should, furthermore, be examined.

The general question on how observed utility measures react to corrections for restriction of range and/or criterion unreliability seem to have received only very limited attention in the literature. The question as to whether corrections for criterion unreliability and/or various forms of restriction of range should be applied in utility analysis had been acknowledged and discussed by several authors [Cascio & Ramos, 1986; Schmidt, Hunter, McKenzie & Muldrow, 1979; Raju, Normand & Burke, 1990; Raju, Burke & Maurer, 1995].

Schmidt, Hunter, McKenzie and Muldrow [1979, p. 612] express their position as follows:

The values of  $r[X,Y]$  and  $SD_Y$  should be those that would hold if applicants were hired randomly with respect to test scores. That is they should be values applicable to the applicant population, the group in which the selection procedure is actually used. Values of  $r[X,Y]$  and  $SD_Y$  computed on incumbents will typically be underestimates because of reduced variance among incumbents on both test and job performance measures." ... Values of  $r_{xy}$  should also be corrected for attenuation due to errors of measurement in the criterion. Random error in the observed measure of job performance causes the test's correlation with that measure to be lower than its correlation with *actual* job performance. Since it is the correlation with actual performance that determines test utility, it is the attenuation-corrected estimate that is needed in the utility formulas.

Raju, Burke and Normand [1990, p. 6], however, disagree with the Schmidt, Hunter, McKenzie and Muldrow [1979] position and maintain that

the psychometric validity of correcting the validity coefficient for unreliability in utility analysis needs to be reassessed.

Raju, Burke and Maurer [1995, pp. 143-144] similarly caution against an over-enthusiastic acceptance of the aforementioned Schmidt, Hunter, McKenzie and Muldrow [1979] stance:

Although several authors have recognized the need for restriction of range corrections in utility analysis, there has been little discussion of the underlying assumptions, and a critical evaluation of the necessity for restriction of range corrections is lacking in the literature.

The objective of future research should be to provide precisely such a critical evaluation by examining the consequences of correcting the various utility estimates for criterion unreliability and/or restriction of range. In contrast to the foregoing citations, however, the analysis should not be restricted to the Brogden-Cronbach-Gleser interpretation of selection utility. Rather than restricting the analysis to a single interpretation of selection utility, the three definitions of payoff suggested by the Taylor-Russell, the Naylor-Shine and the Brogden-Cronbach-Gleser utility models should be considered [Boudreau,

1991; Brogden, 1946; Brogden, 1949; Brogden & Taylor, 1950; Cronbach & Gleser, 1965; Naylor & Shine, 1965; Taylor & Russell, 1993].

Future research should, in addition, examine the effect of corrections for restriction of range and/or criterion unreliability on the evaluation of the investment risk associated with selection procedures.

When viewed from a historical perspective, the evolution of selection utility models present a fairly systematic progression from somewhat unsophisticated models to detailed, complex and rather daunting contemporary models [Boudreau, 1991; Rauschenberger & Schmidt, 1987].

#### 7.2.1.1 Payoff Defined In Terms Of The Validity Coefficient

The utility analysis model with the longest history defines payoff in terms of the validity coefficient. In terms of this classical model, the utility of a [multiple regression] selection strategy is solely a function of the correlation between a weighted, linear composite of predictors and a criterion measure. Selection utility is thus equated with prediction accuracy defined in terms of the residual criterion variance. Over- and underprediction are regarded as equally undesirable, irrespective of the position on the criterion scale where they occur [Boudreau, 1991]. Two translations of the validity coefficient are typically applied to convey its utility implications. The index of forecasting efficiency [Hull, 1928; Kelley, 1923],  $E = [1 - (1 - \rho^2)^{1/2}]$ , indicates the proportional reduction in the standard error of estimate of criterion scores predicted by the regression of the criterion on the weighted linear composite of predictors compared to the standard error of estimate of criterion scores predicted by the criterion mean [Boudreau, 1991; Guilford & Fruchter, 1978; Landy, Farr & Jacobs, 1982]. The coefficient of determination, or the squared validity coefficient, reflects the proportion of variance in the criterion measure accounted for by the weighted linear composite of selection predictors [Boudreau, 1991; Guilford & Fruchter, 1978; Landy, Farr & Jacobs, 1982]. Both these indexes lead to the rather disheartening conclusion that only selection strategies with validities exceeding those normally obtained in validation studies, will have substantial practical utility. The fundamental problem with this line of reasoning, however, lies in its complete disregard of the fact that criterion estimates are not desired as an end in itself, but rather as necessary information required to arrive at a qualitative decision [Cronbach & Gleser, 1965]. Precision in criterion estimation is important in human resource selection, but only in as far as it affects the quality of decision-making.

The classical model's failure to formally acknowledge human resource selection as a form of decision-making, necessarily means that the structural elements characterising selection utility models can only ex post facto be superimposed on it. When forced into a selection utility mold it could be argued,

albeit with some contention, that the classic model implicitly acknowledges decision options in as far as it contrasts prediction based on test scores with chance results, implicitly acknowledges outcomes characterised by the criterion [and therefore the quality attribute in a rudimentary form] and implicitly acknowledges the fact that the quality attribute of the outcome is [partially] a function of the validity coefficient. The classical model, however, fails to, even implicitly, acknowledge the relevance of the quantity and cost outcome attributes, the relevance of the situation in which the selection strategy is used or the existence of a payoff function [total utility is simply equated to validity]. Despite serious deficiencies, the proverbial baby should nevertheless not be thrown out with the bath water. If nothing else, the logic underlying the classic model at least points to the seemingly still to often forgotten fact that, in a human resource selection context, the focus of interest is not the selection instrument, but rather the [evaluated] criterion. Far from being totally abandoned, the classical model in fact forms the bedrock on which all other, more advanced, selection utility models are built. Boudreau [1991, p. 632] attests to this point of view by stating:

Though the deficiencies inherent in these formulas are apparent when viewed from a decision-making perspective, the fundamental notion of expressing the relationship between a predictor and a criterion in terms of the correlation coefficient remains a basic building block of UA models. Later models began to explore ways to embed the correlation coefficient within a set of decision attributes that made it easier to interpret.

#### 7.2.1.2 Payoff Defined In Terms Of The Success Ratio

The Taylor-Russell utility model defines payoff in terms of the success ratio [Taylor & Russell, 1939]. In essence the Taylor-Russell model redefines the prediction problem as a task in predicting a dichotomous criterion created through the formulation of a critical criterion cutoff score, thus reducing the classical bivariate criterion-predictor distribution to a fourfold table [Ghiselli, Campbell & Zedeck, 1981; Taylor & Russell, 1939]. Such redefinition of the prediction problem acknowledges the fact, overlooked by the classical model, that errors in prediction matter to the extent that they cause errors in decisions [Cronbach & Gleser, 1965]. Prediction errors that matter, according to the Taylor-Russell model, are only those where the expected and actual criterion scores fall on opposite sides of the criterion cutoff thus either producing false positive decisions or false negative decisions [Ghiselli, Campbell & Zedeck, 1981; Milkovich & Boudreau, 1994]. The primary error to avoid, according to the Taylor-Russell model, is to select applicants who actuarially should succeed but whose actual performance fall below the minimally acceptable standard [i.e. false positive decisions]. The measure of interest is therefore the success ratio, defined as the proportion of selected applicants who subsequently do succeed as predicted [Boudreau, 1991].



The success ratio [Sv] varies as a function of the proportion of applicants in the applicant group that would succeed on the criterion [i.e.  $Y \geq Y_c$ ] should every applicant be selected [the base rate [BR]], the proportion of applicants in the applicant group actually selected [the selection ratio [SR]] and  $\rho[X,Y]$  [Blum & Naylor, 1968; Milkovich & Boudreau, 1994; Taylor & Russell, 1939].

The model, however, recognises that a high success ratio in itself is not sufficient to establish the practical usefulness of a selection strategy. The Taylor-Russell model equates total utility with the difference in the success ratio afforded by the selection procedure and the success ratio achieved when selecting randomly from the applicant population [i.e. success ratio minus base rate]. The literature generally indicates that the greatest absolute gain in prediction accuracy is achieved under conditions of high predictive validity, low selection ratios and base rates approaching 0.50 [Blum & Naylor, 1968; Boudreau, 1991; Milkovich & Boudreau, 1994]. Boudreau [1991, pp. 632-633], for example contends:

According to the Taylor-Russell tables, when other parameters are held constant, (a) higher validities produce more improved success ratios (because the more linear the relationship, the smaller the area of the distribution lying in the false-positive or false-negative region); (b) lower selection ratios produce more improved success ratios (because lower selection ratios mean more "choosy" selection decisions, and the predictor scores of selectees lie closer to the upper tail of the predictor distribution); and (c) base rate closer to .50 produce more improved success ratios (because as one approaches a base rate of zero, none of the applicants can succeed, thus, selection has less value; as one approaches a base rate of 1.0, all applicants can succeed even without selection, so again, selection has less value).

However, when the behaviour of the Taylor-Russell payoff measure [i.e. the difference in the success ratio afforded by the selection procedure and the success ratio achieved when selecting randomly from the applicant population] is observed graphically over different possible combinations of values of BR, SR and  $\rho[X,Y]$ , such a straightforward conclusion seems unwarranted.

To the extent that the BR, SR and  $\rho[X,Y]$  can be influenced by the human resource function, through changes in recruitment and revision of the selection strategy, it thus becomes possible to actively manage the efficiency of the selection strategy. The Taylor-Russell utility model therefore challenges the pessimism of the classical model on the practical usefulness of selection procedures with validities in the 0.20 - 0.50 range. Taylor and Russell [1939, p. 565-571] articulate their point of view as follow:

The wide-spread acceptance of such measures [as those advocated by the classical model; i.e.  $E$  and  $r'$ ] as the correct way of evaluating correlation coefficients has brought about a considerable pessimism with regard to the validity coefficients which are ordinarily obtainable when tests are tried out in the employment office of a business or industry or in an educational institution. We believe it may be of value to point out the very considerable improvement in selection efficiency which may be obtained with small correlation coefficients.

The Taylor-Russell model formally acknowledges human resource selection as a form of decision-making. It still, however, ignores crucial structural elements of the basic selection utility model. Although the Taylor-Russell model recognises the relevance of the situation in which the selection strategy is used, and acknowledges the quality attribute of outcomes it still ignores both the quantity and cost attributes [Boudreau, 1991; Boudreau, 1989]. The model's measure of quality is, furthermore, bothersome because it firstly does not reflect natural units of value, and secondly, assumes the value of performance to be equal at all points on the criterion scale above the criterion cutoff [Boudreau, 1991; Cascio, 1991b].

### 7.2.1.3 Payoff Defined In Terms Of The Expected Standardised Criterion Score

The Naylor-Shine utility model defines payoff in terms of the expected standardised criterion score of the selected group of applicants on the continuous criterion scale [Blum & Naylor, 1968; Cascio, 1991b; Naylor & Shine, 1965], thereby amending the limitation of the Taylor-Russell utility model of not reflecting the actual range of variation in selectee performance [Boudreau, 1991; Boudreau, 1989].

The Naylor-Shine model assumes a linear, homoscedastic regression of a normally distributed standardised criterion [ $Z_Y$ ] on a normally distributed standardised predictor [ $Z_X$ ]. Under these assumptions, the expected standardised criterion score [ $E[Z_Y]$ ] associated with a specific standardised predictor score would be given by Equation 7.1 [Brogden, 1949; Naylor & Shine, 1965; Cronbach & Gleser, 1965; Cascio, 1991b; Boudreau, 1991]:

$$E[Z_Y] = \rho[X,Y]Z_X \text{-----} 7.1$$

This would then imply that, if the mean standardised predictor score of the selected group of applicants was known, the expected mean standardised criterion performance of the selected group could be found as the product of the validity coefficient and the mean standardised predictor score, as shown in Equation 7.2:

$$E[E[Z_Y]] = \rho[X,Y]E[Z_X] \text{-----} 7.2$$

Brogden [1949] uses Equation 7.2 as the basic point of departure to show that the validity coefficient indicates the ratio of the increase in criterion performance obtained by selecting above a given standard predictor score to the increase that would be obtained by selecting above the same standard score on the criterion itself. Therefore, if an increase of  $\delta$  in mean criterion performance could be obtained by selecting the top  $n\%$  on the criterion, an increase of  $\rho[X,Y]\delta$  would be obtained by selecting the top



n% on the predictor. Given the normality assumptions for the criterion and predictor, and given that strict top-down selection can be assumed, the mean standardised predictor score of the top SR[100]% of the applicant group would equal the product of the selection ratio and the height of the ordinate under the normal distribution at the standardised predictor cutoff [ $\lambda$ ], as shown in Equation 7.3 [Boudreau, 1991; Kelly, 1923; Naylor & Shine, 1965]:

$$E[Z_X] = \lambda/SR \text{-----} 7.3$$

Equation 7.2 can thus be rewritten as shown in Equation 7.4:

$$E[E[Z_Y]] = \rho[X,Y](\lambda/SR) \text{-----} 7.4$$

Equation 7.4 returns the number of standard deviation units the selected group of applicants [assuming strict top-down selection until the desired quota is reached] are expected to perform better on average than a randomly selected group of the same size from the same applicant population. The Naylor-Shine utility model defines total utility in terms of Equation 7.4 as the difference in the mean standardised criterion score between those applicants selected using the selection strategy in question and those selected at random from the applicant population. This, however, has the disadvantage that the unit in terms of which total utility is scaled is ill suited to assess the practical usefulness of a selection strategy in a specific area of application. Transforming, through Equation 7.5, the gain afforded by a selection procedure, expressed in standard deviation units, back to the original, raw score, criterion scale would, to some extent, alleviate the difficulty of interpretation brought about by the standardisation of the criterion scale.

$$E[E[Y|sys]] - E[E[Y|random]] = \rho[X,Y]\sigma[Y](\lambda/SR) \text{-----} 7.5$$

The Naylor-Shine model, like the Taylor-Russell model, formally acknowledges human resource selection as a form of decision-making. It also still ignores crucial structural elements of the basic selection utility model. Although the Naylor-Shine model recognises the relevance of the situation in which the selection strategy is used, and acknowledges the quality attribute of outcomes, it still ignores both the quantity and cost attributes [Boudreau, 1991; Boudreau, 1989]. The model's measure of quality is also troublesome because it still does not reflect natural units of value, although it assumes the value of performance to vary over all points on the criterion scale [Boudreau, 1991; Cascio, 1991b].

#### 7.2.1.4 Payoff Defined In Terms Of A Monetary Valued Criterion

The Brogden-Cronbach-Gleser [B-C-G] selection utility model defines payoff in terms of a monetary valued criterion [Brogden, 1946; Brogden, 1949; Brogden & Taylor, 1950; Cronbach & Gleser, 1965]. The B-C-G selection utility model, similar to the Naylor-Shine model, assumes a linear, homoscedastic regression of a normally distributed criterion [Y], scaled in an appropriate monetary unit, on a normally distributed standardised predictor [ $Z_X$ ]. Under these assumptions, the expected criterion score [ $E[Y]$ ] associated with a specific standardised predictor score would be given by Equation 7.6 [Brogden, 1949; Cascio, 1991b]:

$$E[Y|Z_X] = \alpha + \beta Z_X \text{-----} 7.6$$

Due to the aforementioned distribution assumptions underlying the B-C-G selection utility model, the intercept  $\alpha$  equals the mean criterion score of the applicant group [ $E[Y]$ ] and the regression coefficient  $\beta$  equals the product of the validity coefficient and the ratio of the criterion and predictor standard deviations [Tabachnick & Fidell, 1989]. The standard deviation of the standardised predictor, however, by definition equals unity. Therefore Equation 7.6 can be rewritten as Equation 7.7:

$$E[Y|Z_X] = E[Y] + \rho[X,Y]SD_Y Z_X \text{-----} 7.7$$

If the standardised predictor score fed into Equation 7.7 represents the mean standardised predictor score of the selected applicant group and given  $E[Z_X] = \lambda/SR$ , the absolute monetary value of the mean criterion performance of the selected applicant group would be given by Equation 7.8:

$$\begin{aligned} E[Y|_{\text{sys}}] &= E[Y] + \rho[X,Y]SD_Y E[Z_X] \\ &= E[Y] + \rho[X,Y]SD_Y [\lambda/SR] \text{-----} 7.8 \end{aligned}$$

Equation 7.8 states that the expected value of any new hire in the selected applicant group is the sum of the value that would have resulted from random selection from the applicant pool and the incremental value [i.e. per-selectee value] produced by systematic selection from the same applicant pool [Boudreau, 1991]. The monetary value of the increase in mean job performance afforded by the selection strategy is given by Equation 7.9:

$$\begin{aligned} E[Y|_{\text{sys}}] - E[Y] &= \rho[X,Y]SD_Y E[Z_X] \\ &= \rho[X,Y]SD_Y (\lambda/SR) \text{-----} 7.9 \end{aligned}$$

The marginal or incremental per-selectee utility of the once-off use of a selection procedure with a validity coefficient  $\rho[X,Y]$  to select SR from an applicant group where a one-standard deviation difference in monetary valued performance equals  $SD_Y$  at a per-applicant cost of C is given in Equation 7.10:

$$\Delta U/\text{selectee} = (\rho[X,Y]SD_Y(\lambda/SR)) - (C/SR) \text{-----} 7.10$$

Conversely, the marginal or incremental per-testee utility is given by Equation 7.11 [Cronbach & Gleser, 1965]:

$$\begin{aligned} \Delta U/\text{testee} &= \rho[X,Y]SD_Y\lambda - C \\ &= \Delta U/\text{selectee}[SR] \text{-----} 7.11 \end{aligned}$$

The total gain or utility attained from the once-off use of the selection procedure to select  $N_s$  selectees, representing SR of the applicant group  $[N_a]$ , is given in Equation 7.12 [Brogden, 1949; Cronbach & Gleser, 1965]:

$$\begin{aligned} \Delta U &= N_s\{\rho[X,Y]SD_Y(\lambda/SR) - (C/SR)\} \\ &= \rho[X,Y]SD_Y\lambda N_a - CN_a \text{-----} 7.12 \end{aligned}$$

The two versions of Equation 7.12 are numerically equivalent to each other. The first version of Equation 7.12 reflects the Brogden [1949] argument and is stated in terms of the per-selectee utility multiplied by the number of applicants selected. The second, Cronbach and Gleser [1965], version of Equation 7.12 is derived by multiplying the per-applicant utility by the number of applicants.

A critical component of the B-C-G selection utility model is the standard deviation of monetary valued job performance in the applicant population  $[SD_Y]$ . Accurate measurement of  $SD_Y$  is regarded as a fundamental prerequisite for convincing and credible selection utility analyses. Despite its importance,  $SD_Y$  continues to be a rather elusive quantity to estimate with satisfactory accuracy. The following procedures have been proposed to estimate the standard deviation of the monetary valued criterion [Boudreau, 1989; Boudreau, 1991; Cascio & Ramos, 1986; Cascio, 1991b; Cronbach & Gleser, 1965]:

- cost accounting procedures;
- global estimation procedures;
- individual estimation procedures; and
- proportional rules procedures.

Another, somewhat problematic term contained in the B-C-G selection utility model, is the validity coefficient  $[\rho[X,Y]]$ . Strictly speaking the validity coefficient of interest to Equation 7.12 is the correlation between the predictor and the monetary valued criterion. The monetary valued

performance scores are, however, relatively seldom available and only if a cost accounting or individual estimation procedure is used to estimate  $SD_Y$ . Consequently the term  $\rho[X,Y]$  is interpreted as the correlation between the predictor and a performance scaled criterion measure, under the assumption that a linear, functional relationship exists between performance expressed in behavioral or output units and performance expressed in monetary units. The assumption is thus made that the predictor-criterion correlation is not affected by the unit in terms of which the criterion is scaled [Boudreau, 1991; Hunter & Schmidt, 1982; Schmidt, Hunter, McKenzie & Muldrow, 1979].

Equation 7.12 [and Equation 7.10] shows that the efficiency of a selection strategy improves as the validity coefficient increases, the monetary value of one standard deviation in performance increases, cost of testing per selectee decreases and the selection ratio decreases. Utility is a positive linear function of validity, and for zero cost, is proportional to validity [Brogden, 1946; Brogden, 1949]. For any specific application  $SD_Y$  is a constant which indicates the magnitude and practical significance of individual differences in monetary valued payoff [Cronbach & Gleser, 1965]. Although a decrease in selection ratio increases the term  $\lambda/SR$  [i.e. increases the average ability and performance of the selectees], it at the same time increases the per-selectee cost. Equation 7.12 implies that total utility will only be positive if the predictor cutoff is chosen such that the height of the ordinate under the normal distribution  $[\lambda]$  at the cutoff would exceed the ratio of per-applicant cost to the product of the validity coefficient and the standard deviation of the monetary valued criterion, that is if  $\lambda > C/SD_Y\rho[X,Y]$  [Cronbach & Gleser, 1965]. This would thus imply that for any  $C/SD_Y\rho[X,Y] > 0.3989$  utility must necessarily be negative. For any positive value of  $C/SD_Y\rho[X,Y] < 0.3989$  there exist two standardised predictor scores, symmetrically positioned around the mean, defining a range of lambda values [and therefore cutoff and SR values] which would satisfy the aforementioned expression

Increases in the validity coefficient and  $SD_Y$  would have the effect of pushing these critical cutoff values to more extreme positions, thus permitting profitable selection with more extreme selection ratios. This whole argument logically leads to the conclusion that the lowering of the selection ratio can only benefit utility up to a point [Brogden, 1949]. Cronbach & Gleser [1965] demonstrated, through differentiation of Equation 10 with respect to  $\lambda$ , that the maximum utility per selectee is achieved for the standardised predictor cutoff  $[Z_{Xc}]$  that satisfies Equation 7.13:

$$\lambda - Z_{Xc}[SR] = C/(SD_Y\rho[X,Y]) \text{-----} 7.13$$

If a fixed number of vacancies  $[N_v]$  need to be filled, the standardised predictor cutoff score that satisfies Equation 13, would have to be translated to the required number of applicants to screen  $[N_a]$ ;  $N_a = N_v/SR$  so as to ensure that the optimum selection ratio equals the number of vacancies. If a fixed number of vacancies need to be filled, per-selectee utility, as defined by Equation 10, and total utility as, defined by Equation 7.12, would be maximised by determining the optimum standardised



predictor cutoff and optimum applicant group size through Equation 7.13 [Brogden, 1949; Cronbach & Gleser, 1965]. If, however, no selection quota is in force [i.e. unrestricted selection] and with a fixed number of applicants, Equation 7.12 indicates total utility to be a maximum if the selection ratio is fixed at 0.50 [Cronbach & Gleser, 1965].

Equation 7.12 estimates the utility that would result within the first year from a single application of a valid selection procedure, introduced where previously no procedure, or a totally invalid procedure, had been used, to select a single cohort of selectees from an applicant population. The assumption that the previous procedure can be equated to chance selection may be valid in some cases, but in other situations the new procedure has to compete with a current/alternative procedure. Equation 7.12 can be modified to accommodate this possibility as shown in Equation 7.14 [Cascio, 1991b; Cronbach & Gleser, 1965; Schmidt, Hunter, McKenzie & Muldrow, 1979]:

$$\begin{aligned}\Delta U &= N_s\{(\rho[X,Y]_1 - \rho[X,Y]_2)SD_Y(\lambda/SR) - (C_1 - C_2)/SR\} \\ &= (\rho[X,Y]_1 - \rho[X,Y]_2)SD_Y\lambda N_a - (C_1 - C_2)N_a \text{-----} 7.14\end{aligned}$$

Equation 7.12, furthermore, fails to acknowledge the fact that the benefits that accrue from the selection of any single cohort of selectees are not limited to the first year after selection, but extends as far into the future as the selectees remain in that particular position [Boudreau, 1991]. The costs invested in the selection procedure generate returns to the organisation for as long as those selected remain in the position they were selected into [Cronshaw & Alexander, 1985]. Failing to acknowledge this, Equation 7.12 thus underestimates total utility. Equation 7.12 could, however, be modified to make provision for the projected stream of future returns by incorporating the expected average tenure [T] of the selectees into the equation and by assuming the stream of future returns to remain constant over this time period as shown in Equation 7.15 [Boudreau, 1984; Boudreau, 1989]:

$$\begin{aligned}\Delta U &= N_s\{T\rho[X,Y]SD_Y(\lambda/SR)\} - C/SR\} \\ &= T\rho[X,Y]SD_Y\lambda N_a - CN_a \text{-----} 7.15\end{aligned}$$

When evaluated from a decision-making perspective, the B-C-G selection utility model formally acknowledges all crucial structural elements of the basic selection utility model [Boudreau, 1991; Boudreau, 1989]. Thus all three outcome attributes are explicitly reflected; quantity through mean tenure times number of applicants selected, quality through mean predictor score times validity coefficient times monetary valued standard deviation and cost.

### 7.2.1.5 Refining Payoff Defined In Terms Of A Monetary Valued Criterion

Boudreau [1983a; 1983b; 1989; 1991] pointed out that utility analyses guided by Equation 7.15 would, however, still provide a distorted description of the monetary valued benefits that would result from the implementation of a selection procedure. Such misrepresentation stems from the fact that Equation 7.15 ignores economic considerations normally applied to other institutional financial investment decisions [Boudreau, 1983a; Boudreau, 1983b; Boudreau, 1991; Clark, Hindelang & Pritchard, 1984; Levy & Sarnat, 1994; Lumby, 1994]:

- the tax liability faced by [most] organisations on the returns generated by the investment in valid selection procedures;
- the potential investment returns forfeited on future selection returns;
- the effect of increased criterion performance on variable costs; and
- the effect of employee flows produced by consecutive applications of a selection procedure or the additive cohort effect.

By ignoring the combined effect of variable costs, taxes and the discounting of future earnings, Equation 7.15 effectively stops short of fulfilling its promise of "providing the science of personnel research with a more traditional 'bottom line' interpretation" [Landy, Farr & Jacobs, 1982, p. 38] of institutional benefits comparable to the capital budgeting assessments applied to other, non-personnel, investment options [Boudreau, 1983a; Cronshaw & Alexander, 1985]. Furthermore, by ignoring the combined effects of variable costs, taxes and the discounting of future earnings, Equation 7.15 probably overestimates the total utility of a selection procedure. In contrast, however, by failing to reflect the additive cohorts effect, Equation 7.15 will definitely underestimate the total utility of a selection procedure [Boudreau, 1983b; Boudreau, 1991; Cascio, 1991b].

Boudreau's [1983a] concern with variable costs applies to utility analyses which scale the quality attribute of institutional outcomes in terms of value of output, sales value or revenue. When a selection strategy increases the quantity and/or quality of output which, through its effect on sales value, increases revenue but simultaneously increases/decreases variable costs directly linked to increases affected by the selection procedure [e.g. commission or piece rate remuneration, bonuses, employer pension contributions, raw material costs, scrap or damages], the institutional benefit of such a procedure can not be equated to the increase in sales value. Boudreau [1983a] thus argues that, because a portion of the sacrifices an organisation must concede to produce output will be affected by the effect of the selection procedure on performance, payoff should be defined in terms of the difference between sales revenue and variable service costs or net institutional benefits. Fixed costs, in contrast, do not affect utility estimates [i.e. the change in outcome expressed in terms of revenue or revenue minus fixed costs would be the same], although they would lower the estimate of the mean profit obtained from selected applicants if taken into account [i.e. it would have an effect on a variant of



Equation 5]. Since fixed costs by definition constitute a constant, correcting sales revenue for fixed costs would not affect the monetary scaled standard deviation  $[SD_Y]$ , and therefore not the utility estimate. Boudreau [1983a] proposes two possible avenues through which to include the effect of selection procedures on variable cost in a selection utility model. Because relevant variable costs are those correlating with the increase in sales value produced by the selection procedure, the extent of the variable costs could be estimated by regressing service costs  $[SC]$  on the predictor [i.e.  $E[SC] = [\rho_{X,sc}][\lambda/SR][SD_{sc}]$ ] and subtracting that from the original, sales revenue based, return estimate. Alternatively, by assuming variable costs to be proportional to sales revenue increases, change in net institutional benefits could be obtained by factoring the variable costs component out of increased revenue before adjusting for selection costs. This could be achieved through a parameter  $V$ , where  $V$  equals the ratio of variable costs to sales revenue  $[SV]$ , defining a term  $[1+V]$  which is then multiplicatively combined with the term  $N_s[T][\rho_{X,sv}][\lambda/SR][SD_{sv}]$  before adjusting for selection costs [Boudreau, 1983a; Cascio, 1991b]. The parameter  $V$  becomes negative if the variable costs correlate positively with sales revenue thus reducing the net benefit standard deviation  $[SD_{nb}]$  from  $SD_{sv}$ . Boudreau [1983a] suggests a range of values for the parameter  $V$  from  $-.50$  to  $.33$  which implies an adjustment ranging from  $.50$  to  $1.33$ .

Tax obligations are assessed on an organisation's reported profit. To the extent that improved selection procedures contribute to increased profits it will also contribute to increased tax liability. Utility estimates ignoring the impact of improved selection on tax liabilities would therefore overstate the net institutional benefits derived from improved selection procedures. Because taxes are assessed on reported profits, adjusting for taxes produces a proportional reduction in both revenue and selection costs [Boudreau, 1983a]. The appropriate tax adjustment, given the preceding argument, is the tax rate  $[TAX]$  applicable to the increase in reported profits attributable to improved selection [i.e. the marginal tax rate]. The total utility estimate, after adjustment for variable costs, should thus be multiplied with  $[1-TAX]$ .

Selection procedures produce returns which accrue to the organisation over time. Equation 7.15 acknowledges this fact, but ignores the fact that returns received in different time periods are subject to different opportunity costs [Boudreau, 1983a; Boudreau, 1991]. By simply adding the returns obtained in future years to the returns obtained during the initial years, Equation 7.15 wrongfully equates future monetary values with present monetary value thus ignoring the fact that immediately received returns [and deferred costs] would/could be invested to earn interest [Boudreau, 1983a; Boudreau, 1991]. Returns received at different points in time should, therefore, first be converted to a common point in time before accumulating the returns [Clark, Hindelang & Pritchard, 1984; Levy & Sarnat, 1994; Lumby, 1994]. The present time constitutes such a common point. The net present value  $[NPV]$  of a constant yearly stream of net benefits  $[A_t]$  generated over  $T$  years by investing in an improved selection

procedure applied once to select a single cohort of  $N_s$  selectees, given a constant discount or interest rate of  $i$ , is shown in Equation 7.16 [Clark, Hindelang & Pritchard, 1984; Lumby, 1994;]:

$$NPV = N_s \{ \sum [A_t / (1+i)^t] \} \text{-----} 7.16$$

Building the effect of discounting into Equation 7.15 would require an estimate of the appropriate discount rate. The discount rate in general flows from the implicit alternative to investing in a selection procedure [or any other human resource intervention], namely lending the investment amount out on the capital market at the market rate of interest [Lumby, 1994]. According to Boudreau [1983a, p. 566]:

the appropriate discount rate for utility analysis should be the rate applied to uninflated benefits and costs given the organisation's evaluation of overall risk and return requirements. .... Because personnel programs will be employed in all types of organisations from very risky endeavors to relatively risk-free endeavors, the discount rate [ $i$ ] must reflect the fact that risky firms must earn a risk premium and face a higher discount rate than less risky firms.

Combining the effects of variable costs, taxes and the discounting of future earnings in Equation 7.15 produces Equation 7.17:

$$\Delta U = N_s \{ \sum [1 / (1+i)^t] \} SD_{sv} (1+V) (1-TAX) \rho_{X,sv} (\lambda/SR) - C(1-TAX) \text{-----} 7.17$$

Equation 7.17 reflects the fact that the costs of a selection procedure occur only at the stage of selection and thus are not subject to discounting [Boudreau, 1983a]. Furthermore, Equation 7.17 assumes all parameters to remain constant over the  $T$  time periods. This assumption could readily be relaxed by converting all relevant parameters that could be subject to change to variables, but that would create the practical problem of estimating an appropriate array of values for each variable over the  $T$  time periods.

Cronshaw and Alexander [1985] support the Boudreau [1983a] position that investments in human resource interventions must be evaluated similarly to other investment options. Contrary to Boudreau [1983a, 1983b], however, Cronshaw & Alexander [1985] do not propose any significant changes to the basic B-C-G selection utility model [Equations 7.12 & 7.15], but rather suggest a number of capital budgeting methods to follow on a point estimate of the expected return to the selection investment. The following capital budgeting investment appraisal methods [Clark, Hindelang & Pritchard, 1984; Du Plessis, 1986; Levy & Sarnat, 1994; Lumby, 1994] represent conceptual tools to integrate selection utility into the broader financial decision-making processes within organisations:

- pay-back period [PP];
- return on investment [ROI];

- net present value [NPV];
- profitability index [PI]; and
- internal rate of return [IRR].

The pay-back period [PP] is defined as the number of years an organisation requires to recover its original investment from returns [Clark, Hindelang & Pritchard, 1984; Cronshaw & Alexander, 1985; Lumby, 1994]. In selection investment applications, the pay-back period becomes especially relevant for costly selection procedures applied to select a relatively small number of new hires into high turnover jobs. Under these conditions there exists the real possibility that the pay-back period might approach or exceed the expected tenure of the new cohort of hires; the selection procedure may therefore not yield any real economic benefit to the organisation. Return on investment [ROI] compares annual returns with the investment in the selection project. Net present value [NPV] is defined as the difference between the discounted sum of the projected future returns and the total amount invested. A projects NPV represents the increase/decrease in return [i.e. excess return] that would result from investing in the selection procedure rather than lending the investment amount on the capital market at the market rate of interest. Consequently investment in any selection procedure with a zero or positive NPV would be acceptable. A negative NPV project would be unacceptable because it would make a loss relative to a capital market investment, it would produce a return less than that available on the capital market and it would not generate sufficient returns to repay the financial costs of undertaking it [Lumby, 1994]. The profitability index [PI] is defined as the ratio of the present value of cash inflows to cash outflows. A ratio of one or greater, indicates that the yield of investing in an improved selection procedure is equivalent to or greater than the discount rate [Clark, Hindelang & Pritchard, 1984]. The internal rate of return [IRR] is defined as the rate which equates the net present value of cash inflows with cash outflows. The IRR can thus be defined as the rate of discount which, when applied to the selection investment's cash flow, would produce a zero NPV [Lumby, 1994; Clark, Hindelang & Pritchard, 1984]. Only projects with an IRR greater than the market rate of interest [or any other cut-off rate] will be accepted [Lumby, 1994]. According to Cronshaw & Alexander [1985] the aforementioned capital budgeting investment appraisal methods could be represented symbolically, in terms of the Equation 7.15 version of the B-C-G selection utility model, as shown by Equations 7.18-7.22:

$$PP = (C_1 + C_0) / \{N_s \rho[X, Y] SD_Y(\lambda/SR)\} \text{-----} 7.18$$

$$ROI = \{ \{N_s \rho[X, Y] SD_Y(\lambda/SR)\} / (C_1 + C_0) \} 100 \text{-----} 7.19$$

$$NPV = \Sigma \{ \{N_s \rho[X, Y] SD_Y(\lambda/SR)\}_t / [1+k]^t \} - (C_1 + C_0) \text{-----} 7.20$$

$$PI = \{ \Sigma \{ \{N_s \rho[X, Y] SD_Y(\lambda/SR)\}_t / (1+k)^t \} / (C_1 + C_0) \} \text{-----} 7.21$$

$$IRR = j; \Sigma \{ \{N_s \rho[X, Y] SD_Y(\lambda/SR)\}_t / (1+j)^t \} = (C_1 + C_0) \text{-----} 7.22$$



Boudreau [1991] disagrees with the Cronshaw and Alexander [1985] formulations and suggests that more consistent formulations would result if total returns [R] would be corrected by the implementation costs [C<sub>1</sub>] before comparing it to the original costs [C<sub>0</sub>] of developing and validating the selection procedure.

Equation 7.17, like all the B-C-G utility models preceding it, estimates total utility as the increase in average monetary valued payoff resulting from adding a single cohort of selectees to the existing work force through the use of a battery of predictors instead of random selection from the applicant population. According to Boudreau [1983b; 1991], considering only a single cohort, however, unnecessarily limits these utility models. Selection procedures are continuously reapplied as the organisation expands and/or employees move out of, or elsewhere in, the organisation. Organisations do not invest in a selection procedure to use it once and then to discontinue its use, but rather to continuously reapply the procedure to admit new members into the work force [Boudreau, 1991]. Thus, Boudreau [1983b] argues that the decision on the practical usefulness of a selection procedure cannot be based on the results of a utility analysis focusing on a single application only. Equation 7.17, and its B-C-G utility model predecessors, reflect the quantity attribute by multiplying the number hired with the mean tenure of selectees [T]. Boudreau's [1983b] employee flows utility model, in contrast, reflects quantity of employee-years of output in terms of the number of treated employees in the work force [N<sub>k</sub>], k time period in the future. The term N<sub>k</sub> represents the end result of the combined effect of the number of selected employees entering the work force [N<sub>at</sub>], the number of previously selected employees leaving the work force [N<sub>st</sub>] and the average tenure of employees selected into the organisation [i.e. N<sub>k</sub> = Σ[N<sub>at</sub>-N<sub>st</sub>]; t=1, 2, ..., k]. For any given time period t=q, 1 < q < k, utility is still reflected by the product of quantity [N<sub>q</sub>] times quality [ρ[X,Y][SD<sub>Y</sub>][λ/SR]] minus cost. To obtain a total utility estimate these separate time period utility estimates are summed over all future time periods [k=1, 2, ..., F]. Boudreau's [1983b; 1991] employee acquisition flows utility model is formally explicated in Equation 7.23:

$$\Delta U = \frac{\Sigma(\Sigma(N_{at}-N_{st}))\{(1/(1+i)^k)SD_{sv}(1+V)(1-TAX)r_{x,sv}(\lambda/SR)\} - \Sigma\{C_k(1-TAX)\{1/(1+i)^{(k-1)}\}\}}{\Sigma\{C_k(1-TAX)\{1/(1+i)^{(k-1)}\}\}} \quad 7.23$$

The classic B-C-G utility formulations [Equation 7.12] have been modified to apply to performance enhancing human resource interventions [e.g. training or performance feedback] following on selection [Florin-Thuma & Boudreau, 1987; Landy, Farr & Jacobs, 1982; Mathieu & Leonard, 1987; Schmidt, Hunter & Pearlman, 1982] by substituting the product [SD<sub>Y</sub>][ρ[X,Y]] reflecting the difference in average criterion performance between systematically and randomly selected applicants expressed in standard deviation units, with d, the difference in criterion performance between the treated and untreated groups, expressed in standard deviation units. All subsequent developments on the basic B-C-G formulation [Equations 7.14, 7.15, 7.17 & 7.23] apply equally well, in their modified form, to

human resource interventions aimed at improving employee performance in their existing assignments [Milkovich & Boudreau, 1994].

A noticeable shortcoming in the utility analysis approaches discussed thus far, is their inability to account for human resource management activities preceding selection [i.e. recruitment] and employee separation [i.e. layoffs, retirement, quits & dismissals]. Boudreau and Rynes [1985], and Boudreau and Berger [1985] addressed this shortcoming by proposing modifications of Equation 7.23 that would integrate the effects of recruitment and selection into a single decision-theoretic utility model [Boudreau, 1991; Boudreau & Rynes, 1985] and by developing a utility model that could encompass not only the effects of employee acquisitions but also of employee separations [Boudreau, 1991; Boudreau & Berger, 1985a].

All versions of the basic B-C-G selection utility model discussed thus far, scaled utility as the difference from an unknown utility level that would have resulted if the selection strategy would not have been used. All versions of the basic B-C-G selection utility model discussed thus far, also assumed that all selection strategy options would be implemented within the same applicant population [Boudreau, 1991]. The latter assumption implies that all selection options considered are teamed up with the same recruitment strategy. If this would not be the case, any one of the B-C-G selection utility models discussed thus far, could provide a misleading indication of the most beneficial selection option because it would ignore the effects of the recruitment strategy it is teamed up with on the applicant pool [Boudreau & Rynes, 1985; Boudreau, 1991]. The recruitment strategy, specifically the recruitment method, the nature of the recruitment message, the nature and level of required applicant qualifications and the nature of administrative procedures [Boudreau & Rynes, 1985], could influence the size of the applicant pool as well as the parameters of the marginal criterion and predictor distributions in the applicant population. The recruitment strategy could therefore influence all parameters of the B-C-G selection utility models discussed thus far. Thus Boudreau and Rynes [1985] argue that a particular recruitment strategy-selection strategy combination could produce the greatest total value, although the selection procedure of choice may offer a lower incremental utility relative to other selection options, if the recruitment strategy produces a substantial increase in the average applicant value. To express utility in absolute terms Boudreau and Rynes [1985] essentially add the incremental utility afforded by the selection option under consideration to the expected value of the discounted, after cost, after tax monetary valued payoff resulting from random selection from the recruited applicant pool. Boudreau and Rynes [1985] proposed three changes to Equation 7.23 to reflect the combined effect of recruitment and selection on utility. They firstly introduced a variable  $C_{rk}$  to represent the cost associated with a particular recruitment strategy  $[r]$  that generates the applicant pool from which  $N_{ak}$  new hires are selected in time period  $k$ . They secondly introduced a term  $[\mu_{svr} - \mu_{scr}]$  to represent the average after-cost service/sales value of the recruited applicant pool. The third modification proposed by Boudreau and Rynes [1985] was to formally change the status of  $\rho[X,Y]$ ,  $\lambda/SR$ ,  $SD_Y$  and  $C$  from

constants to variables with values contingent on the recruitment strategy employed. The expected value of the discounted, after-cost, after tax monetary valued payoff from using selection procedure  $x$  under recruitment strategy  $r$ , evaluated for  $F$  future periods is given by Equation 7.24:

$$E[U]_{X,r} = \left\{ \sum (1/1+i)^k \left\{ \sum (N_{at}-N_{st})(\rho_{X,svr})(Z_{Xr})(SD_{svr})(1+V) \right\} + \right. \\ \left. \sum (1/1+i)^k \left\{ \sum (N_{atr}-N_{str})(\mu_{svr}-\mu_{scr}) \right\} - \right. \\ \left. \sum (1/1+i)^{(k-1)}(C_{sk}+C_{rk}) \right\} (1-TAX) \text{-----} 7.24$$

Significant similarity exist between employee acquisitions through selection procedures and employee separations through quits, layoffs and dismissals [Boudreau, 1991]. Both processes involve employee movement across the organisational boundary [Boudreau & Berger, 1985a]. Furthermore, both processes in essence refer to a procedure whereby a subset of employees is selected from a larger "applicant" pool. Human resource selection procedures acquire subsets of employees from a larger applicant pool outside the organisation while human resource separation procedures select subsets of employees from a larger pool of current employees to remain in the organisation [Boudreau & Berger, 1985a; Milkovich & Boudreau, 1994]. Boudreau and Berger [1985] exploited these similarities between employee acquisitions and employee separations to develop an external employee movement utility model that could encompass the effects of both processes. The multiple-cohort acquisition and retention utility model [Boudreau & Berger, 1985a] estimates the absolute utility of the work force, rather than the incremental utility added by, procedures designed to regulate employee movement across organisational boundaries, under any one of the following three possible scenarios:

- repeated acquisitions without separations over time [i.e. the work force is systematically increased through selection];
- repeated unreplaced separations over time [i.e. the work force is systematically reduced through separations]; and
- repeated separations and acquisitions over time [i.e. the work force is either maintained, increased or reduced depending on the number of employees entering and leaving during each time period].

The multiple-cohort acquisition and retention utility model [Boudreau & Berger, 1985a] essentially integrates two distinct utility models; a repeated selection utility [or multiple cohort acquisition] model estimating the discounted, after-cost, after-tax work force value that would result in a pure growth situation and a repeated separation [or multiple cohort retention] utility model estimating the discounted, after-cost, after-tax work force value that would result in a pure shrinkage situation. The multiple cohort acquisition utility model estimates total work force value by firstly estimating the work force value at the beginning of the analysis as the number of incumbent employees  $[N_{i0}]$  in the period prior to the implementation of the selection procedure times their quality [i.e. their mean service value minus the mean service costs associated with that force  $[sv_0-sc_0]$ ]. The implementation of a selection



procedure increases this initial work force value by the level of utility expected from random selection from the applicant pool [i.e. the number of new hires acquired  $[N_{at}]$  times the difference between the mean service value  $[\mu_{svat}]$  and mean service cost  $[\mu_{scat}]$  associated with applicants of the applicant population] plus the incremental utility offered by the valid selection procedure [i.e. the number of new hires acquired  $[N_{at}]$  times  $[r_{x,sv}][\lambda/SR][SD_{sv}]$ ]. Finally, the transaction costs associated with adding average service value employees to the work force  $[C_a]$  and the incremental transaction costs associated with systematic selection  $[\Delta C_a]$  are subtracted from the sum of the aforementioned three estimates [Boudreau & Berger, 1985a; Boudreau, 1991]. The multiple cohort retention utility model estimates the total work force value that would result in a pure shrinkage situation by also firstly estimating the work force value at the beginning of the analysis as the number of incumbent employees  $[N_{i0}]$  in the period prior to the implementation of the separation procedure times their quality [i.e. their mean service value minus the mean service costs associated with that force  $[sv_0 - sc_0]$ ]. If separations occur randomly [i.e. the decision criterion correlates zero with service value] the total work force utility associated with the remaining work force is reduced from the prior value by the number of employees separating  $[N_{st}]$  times the difference in average service value and average service cost  $[sv_0 - sc_0]$  associated with the prior incumbent employee population [i.e. random separations would not effect the average service value and average service costs but would reduce the size of the work force]. On the other hand, if separations/retentions do not occur randomly, and if the criterion  $[q]$  on which such separation decisions are based, correlates with service value  $[r_{q,sv(t-1)}]$ , the prior work force value will have to be adjusted for the effect if retentions would have occurred randomly and for the incremental effect attributable to the systematic retention procedure [Boudreau & Berger, 1985a]. The latter effect is estimated through essentially the same logic underlying the earlier B-C-G selection utility models. Assuming the retention criterion  $[q]$  to be positively and linearly related to service value, assuming thus a strict bottom-up separation policy, and assuming  $q$  to be normally distributed in the pre-separation work force, the mean retention criterion score of the retained employees  $[Z_{qR}]$  would equal the ratio of the height of the ordinate under the standardised normal distribution at the  $q$ -cutoff to the retention ratio [i.e. the number of employees retained divided by the pre-separation number of employees]. The multiple-cohort acquisition and retention utility model [Boudreau & Berger, 1985a], shown in Equation 7.25, represents an integration of these two selection utility models in a comprehensive external employee movement utility model:

$$U_w = \{ \Sigma(1/1+i)^k [(N_{i0})(sv_0 - sc_0) + \Sigma(N_{at})(\mu_{svat} - \mu_{scat}) + \Sigma(N_{at})(r_{x,svat})(\lambda/SR)(SD_{svat})(1 + V_t) - \Sigma(N_{st})(sv_{i(t-1)} - sc_{i(t-1)}) + \Sigma(N_{i(t-1)} - N_{st})(d_{sv})(SD_{sv})(1 + V_t) - (1/1+i)^{(k-1)}(C_{ak} + \Delta C_{ak} + C_{sk} + \Delta C_{sk}) \} (1 - TAX) \} \quad 7.25$$

The multiple-cohort acquisition and retention utility model [Boudreau & Berger, 1985a], although substantially more comprehensive and complicated than the basic B-C-G selection utility model it

evolved from, still has the limitation that it only focuses on external employee movement across organisational boundaries while ignoring internal employee movement between jobs within the organisation [Boudreau, 1991; Boudreau & Berger, 1985b]. Internal employee movement is, however, important not only because internal movement affects the work force utility of jobs that internally acquire employees but also because such movement affects the work force utility of the jobs that internally separate employees [Milkovich & Boudreau, 1994; Boudreau & Berger, 1985b]. Internal selection procedures which extract the superior employees from lower level jobs can have severe organisational consequences especially if the promotion criterion shows little if any correlation with performance in the subsequent position. The internal/external employee movement selection utility model represents an integration of the Boudreau and Berger [1985a] external employee movement utility model and an adaptation thereof to reflect the effect of internal employee flows. Internal employee movement involves a separation from one organisational job and an acquisition by another. The concepts of retention and selection utility, as applied by the Boudreau and Berger [1985a] external employee movement utility model, are thus also applicable to the analysis of internal employee movement. A single internal movement, however, in contrast to any single boundary crossing movement, affects both retention and selection utility [Boudreau, 1991; Boudreau & Berger, 1985b]. Total work force utility is the discounted, after-tax, after-cost sum of the work force value in the jobs contained in the organisation structure. The work force value in any specific job, in turn, at time period  $t=1$  depends on the quantity and quality of the initial workforce, the quantity and quality of employees flowing in and out of the job due to internal and external separations and acquisitions. No formal algebraic expression of the integrated internal/external employee movement utility model could be located in the literature.

#### 7.2.1.6 Analysis Of Selection In Terms Of Risk

In the modern idiom the term risk refers to a hazard or chance of harm, injury or loss. Modern investment analysis, however, interprets risk in terms of its original Latin [*risicum*] meaning as chance deviations [both positive and negative] from the expected return [Levy & Sarnat, 1994]. The term risk thus applies to an investment option whose actual return is not known in advance with absolute certainty, but for which a distribution of alternative payoffs and their associated probabilities are known. The term risk refers to the amount of variability or dispersion present in the probability distribution of returns associated with the decision option [Clark, Hindelang & Pritchard, 1984; Levy & Sarnat, 1994]. All other considerations kept constant, the option showing the least risk [i.e. the least variability in possible returns] would be preferred.

All B-C-G selection utility models discussed thus far estimate the expected absolute or incremental value of the returns obtained from investing in human resource selection procedures. The utility estimate, however, depends on several parameter estimates which are uncertain. A probability distribution of estimated utility derived from investing in a human resource selection procedure would therefore result if different parameter estimates would be selected from the probability distribution of each individual parameter. The variance associated with utility estimates has, however, been largely ignored [Alexander & Barrick, 1987; Boudreau, 1991; Rich & Boudreau, 1987]. The following methods could be used to assess risk attributable to utility variability [Rich & Boudreau, 1987]:

- sensitivity analysis;
- break-even analysis;
- algebraic variability estimation; and
- Monte Carlo analysis.

In a sensitivity analysis each of the utility parameters are varied from a low value through the actual estimate to a high value while holding all other utility parameter estimates constant. The resulting array of utility estimates is subsequently examined to determine the effect of parameter estimate variability on the total estimate [Rich & Boudreau, 1987]. Break-even analysis represents an extension of the logic underlying sensitivity analysis [Boudreau, 1984; Rich & Boudreau, 1987]. Break-even analysis determines the individual parameter values at which the human resource programme's benefits are equal to [i.e. even with] the programme's cost. If the probability of obtaining an actual parameter estimate equal to, or less beneficial than, the break-even value, is sufficiently large, then relatively little risk is involved in rejecting the programme. Conversely, if the probability of obtaining an actual parameter estimate more beneficial than the break-even value is sufficiently large, then relatively little risk is involved in accepting the programme [Boudreau, 1984; Boudreau, 1991; Rich & Boudreau, 1987]. Boudreau [1984] points out that under these conditions the appropriate decision regarding the acceptance or rejection of the programme can be made without resorting to detailed and complex parameter estimation procedures. Algebraic variability estimation utilises the fact that the cost and return components of the B-C-G selection utility equation can both be written in the basic form  $Y = \beta X$ , where  $X$  represents a multiplicative combination of parameter estimates [i.e.  $X = [X_1 \times X_2 \times X_3 \times \dots \times X_k]$ ] so that  $\sigma^2[Y] = \beta^2 \sigma^2[X]$ . The term  $\sigma^2[X]$  in turn would be calculated through Equation 7.26 [Alexander & Barrick, 1987]:

$$\sigma^2[X] = \prod [\mu_i^2 + \sigma_i^2(n_i - 1)/n_i] - \prod [\mu_i^2 - \sigma_i^2/n_i]; i = 1, 2, \dots, k \text{ ----- } 7.26$$

Assuming the cost and return components of the B-C-G selection utility models to be independent, the variance in the overall utility estimate is simply the sum of  $\sigma^2_r$  and  $\sigma^2_c$  [Alexander & Barrick, 1987]. Monte Carlo simulation analysis requires a description of each utility model parameter in terms of its expected value, variance and distribution shape. Through a computer driven iterative procedure of

randomly choosing values from the parameter distributions and subsequently calculating overall utility, a distribution of utility values is obtained. The statistical characteristics of the obtained utility distribution could then be derived.

### **7.2.2 Effect Of Corrections For Random Measurement Error And Nonrandom Selection On Fairness Assessments Of Selection Procedures**

Human resource selection constitutes a potent instrument enabling the human resource function to add value to the organisation by virtue of its ability to regulate the quality and quantity of employees flowing into, through and out of the organisation. Human resource selection procedures derive their ability to add value to the organisation from their capability to discriminate between applicants in terms of attributes relevant to job performance. Selection measures are designed to discriminate and in order to accomplish its professed objective it must do so [Cascio, 1991a]. However, due to the relative visibility of the selection mechanism's regulatory effect on the access to employment opportunities, the question readily arises whether the selection strategy discriminates fairly. Selection fairness, however, represents an exceedingly elusive concept to pin down with a definitive constitutive definition. The problem is firstly that the concept cannot be adequately defined purely in terms of psychometric considerations without any attention to moral/ethical considerations. The inescapable fact is that, due to differences in values, one man's foul is another man's fair. The problem is further complicated by the fact that a number of different definitions and models of fairness exist which differ in terms of their implicit ethical positions and which, under certain conditions, are contradictory in terms of their assessment of the fairness of a selection strategy and their recommendations on remedial action [Arvey & Faley, 1988; Cascio, 1991a; Petersen & Novick, 1976]. Three distinct fundamental ethical positions [Hunter & Schmidt, 1976] were previously identified. A fairness model, based on one of these ethical positions [or a variant thereof], serves the purpose of formalising the interpretation of the fairness concept and thus permitting the deduction of a formal investigative procedure needed to assess the fairness of a particular selection strategy should such a strategy be challenged to disprove a *prima facie* showing of adverse impact [Arvey & Faley, 1988; Singer, 1993]. A variety of fairness models have been proposed [Arvey & Faley, 1988; Cascio, 1991a; Petersen & Novick, 1976]. These fairness models examine the effect of the selection decision function on different subgroups contained in the applicant population by simulating the selection process on a representative sample from the applicant population. At least thirteen different selection fairness models can be distinguished [Arvey & Faley, 1988; Cascio, 1991a; Petersen & Novick, 1976].

Since it had been shown that statistical corrections for criterion unreliability and/or restriction of range can affect the magnitude of the validity coefficient, the probability of observing the validity coefficient



under  $H_0$ , the parameters of the regression of the criterion on the predictor and the variance of the conditional criterion distribution, subsequent research should examine how these effects impact on the judgments delivered on the fairness of selection procedures by various selection fairness models. The fundamental objective of future research in this regard should therefore be to determine whether the finding/verdict of the various fairness models when applied to a specific selection strategy and their recommendations on remedial action, if required, would change as a result of these corrections.

At least three of the aforementioned array of fairness models [Petersen & Novick, 1976] should be formally examined in terms of their reaction to statistical corrections for restriction of range and/or criterion unreliability, namely:

- the regression or Cleary model;
- the equal risk or Einhorn-Bass model; and
- the constant ratio or Thorndike model.

#### 7.2.2.1 The Cleary Model Of Selection Fairness

The regression or Cleary model of selection fairness defines fairness in terms of differences in regression slopes and/or intercepts across the subgroups [ $\pi_1$  and  $\pi_2$ ] comprising the applicant population [Arvey & Faley, 1988; ; Cascio, 1991a; Maxwell & Arvey, 1993; Petersen & Novick, 1976]. According to Cleary [1968, p. 115]:

A test is biased for members of a subgroup of the population if, in the prediction of the criterion for which the test was designed, consistent nonzero errors of prediction are made for members of the subgroup. In other words, the test is biased if the criterion score predicted from the common regression line is consistently too high or too low for members of the subgroup. With this definition of bias, there may be a connotation of unfair, particularly if the use of the test produces a prediction that is too low. If the test is used for selection, members of a subgroup may be rejected when they were capable of adequate performance.

The Cleary model examines the fairness of a selection strategy by fitting a saturated regression equation, exemplify by Equation 7.28 shown below, and testing the hypothesis  $H_{01}: \beta_2 = \beta_3 = 0$  against the alternative hypothesis  $H_a$ : at least one of the two parameters is not zero [Berenson, Levine & Goldstein, 1983; Kleinbaum & Kupper, 1978]. Should  $H_{01}$  not be rejected it would imply the use of the common regression equation to be fair. Should  $H_{01}$ , however, be rejected it would imply the use of the common regression equation to be unfair because it would imply the separate regression equations to differ in terms of slope and/or intercept [i.e. one would have to conclude that the regression models fitted to the two subgroups are not coincident].

$$E[Y|X] = \alpha + \beta_1 X + \beta_2 D + \beta_3 X \cdot D \text{-----} 7.28$$

Where D denotes a dummy variable, coded such that:

D = 0 for group  $\pi_1$

D = 1 for group  $\pi_2$

The appropriate remedy, should the latter situation prevail, is contingent on the explanation for the rejection of  $H_0$ . The Cleary model's prescription for a diagnosed unfair selection strategy thus depends on whether there exists an equivalent incremental difference in criterion performance across applicants from the two subgroups, regardless of predictor performance [i.e.  $\beta_2 \neq 0$ , but the interaction parameter  $\beta_3$  can be assumed zero] or a non-equivalent incremental difference in criterion performance across applicants from the two subgroups, dependent on the ability level of the applicants [i.e. there exists a subgroup x predictor performance interaction effect on criterion performance;  $\beta_3 \neq 0$ ] [Berenson, Levine & Goldstein, 1983; Kleinbaum & Kupper, 1978]. The Cleary solution to the fairness problem thus dictates that the information category entries in the strategy matrix [Cronbach & Gleser, 1965] should be derived from an appropriately expanded multiple regression equation containing the grouping variable as a main effect and/or in interaction with the predictor.

#### 7.2.2.2 The Einhorn-Bass Model Of Selection Fairness

The equal risk or Einhorn-Bass selection fairness model operationalises the concept in terms of differences in the probability of success conditional on predictor performance [i.e. a selection strategy would be considered unfair if  $P[Y \geq Y_c | X_i, \pi_1] \neq P[Y \geq Y_c | X_i, \pi_2]$  for all  $i$ ] [Cascio, 1991a; Einhorn & Bass, 1971; Petersen & Novick, 1976]. The Einhorn-Bass conceptualisation thus corresponds to the Guion [1966, p. 26] definition of unfair discrimination:

Unfair discrimination exists when persons with equal probabilities of success on the job have unequal probabilities of being hired for the job.

The equal risk model would therefore judge any selection strategy unfair should it be considered unfair by the Cleary model. In addition, however, it would also consider the selection strategy unfair if the criterion variance conditional on predictor performance differs across the two applicant subgroups [i.e.  $\sigma^2[Y|X; \pi_1] \neq \sigma^2[Y|X; \pi_2]$ ] [Cascio, 1991a; Einhorn & Bass, 1971; Petersen & Novick, 1976]. The Einhorn-Bass solution to the fairness problem thus dictates that the information category entries [i.e.  $P[Y \geq Y_c | X_i, \pi_j]$ ] in the strategy matrix [Cronbach & Gleser, 1965] should be obtained by deriving



$E[Y|X_i;\pi_j]$  from the appropriate regression equation and subsequently, transforming  $Y_c$  to a standard score in the conditional criterion distribution [assuming normality] by using the appropriate standard error of estimate as denominator [Berenson, Levine & Goldstein, 1983; Einhorn & Bass, 1971; Kleinbaum & Kupper, 1978].

### 7.2.2.3 The Thorndike Model Of Selection Fairness

The constant ratio or Thorndike model of selection fairness considers a selection strategy fair if it selects the same proportion of applicants from the various applicant subgroups that would qualify for selection on the basis of the criterion [Cascio, 1991a; Petersen & Novick, 1976; Thorndike, 1971]. Thorndike [1971, p. 63] approaches the question of selection fairness from a group perspective when he proposes that in a fair selection strategy:

the qualifying scores on a test should be set at levels that will qualify applicants in the two groups in proportion to the fraction of the two groups reaching a specified level of criterion performance.

The most generic interpretation of the constant ratio model would be to consider a selection strategy fair if the ratio of the selection ratio to the base rate would be equal across the subgroups of the applicant population [Petersen & Novick, 1976]. Should a selection strategy be judged unfair towards a specific subgroup the solution to the fairness problem, according to the constant ratio model, would be to adjust the predictor cutoff scores [Thorndike, 1971].

### 7.2.2 Additional Specific Recommendations For Further Research

In addition to examining the effect of corrections for restriction of range and/or criterion unreliability on utility and fairness assessments, future research should also examine:

- the effect of corrections for restriction of range and/or criterion unreliability on the standard errors of the regression coefficients and standard error of estimate of the regression of Y on X;
- the effect of joint corrections for Case C restriction of range and criterion unreliability on the validity coefficient, the standard error of the validity coefficient, the empirical exceedence probability, the slope and intercept parameters of the regression of Y on X and the standard error of estimate;

- the correlation  $\rho[E(T_Y|X), T_Y]$ , relative to  $\rho[X, Y]$ ,  $\rho[X, T_Y]$ , and  $\rho[T_X, T_Y]$  and its standard error;
- the effect of Case 2 [Case A] restriction of range on Y on the parameters of the regression of Y on X.

Future research should also extend the analyses in chapter 6 to the multiple regression model. The effect of corrections for restriction of range and/or criterion unreliability on selection decision-making should also be examined from the perspective of alternative selection strategies.

## REFERENCES

- Alexander, R.A. [1988]. Group homogeneity, range restriction, and range enhancement effects on correlations. *Personnel Psychology*, 41, 773-777.
- Alexander, R.A. & Barrick, M.R. [1987]. Estimating the standard error of projected dollar gains in utility analysis. *Journal of Applied Psychology*, 72 [3], 475-479.
- Allan, N.L. & Dunbar, S.B. [1990]. Standard errors of correlations adjusted for incidental selection. *Applied Psychological Measurement*, 14 [1], 83-94.
- American Psychological Association. [1954]. Technical recommendations for psychological tests and diagnostic techniques. *Psychological Bulletin*, 51 [2,Pt. 2].
- American Psychological Association. [1966]. *Standards for educational and psychological tests and manuals*. Washington, DC: Author.
- American Psychological Association, American Educational Research Association & National Council on Measurement in Education. [1974]. *Standards for educational and psychological tests*. Washington; American Psychological Association.
- American Psychological Association, American Educational Research Association & National Council on Measurement in Education. [1985]. *Standards for educational and psychological tests*. Washington; American Psychological Association.
- Angoff, W.H. [1953]. Test reliability and effective test length. *Psychometrika*, 18 [1], 1-14.
- Arvey, R.D. & Faley, R.H. [1988]. *Fairness in selecting employees* [Second edition]. Reading, Mass.: Addison-Wesley.
- Austin, J.T. & Villanova, P. [1992]. The criterion problem: 1917-1992. *Journal of Applied Psychology*, 77 [6], 836-874.
- Babbel, D.F., Stricker, R. & Vanderhoof, I.T. [1994]. A modern approach to performance measurement for insurers. In R.L. D'Ecclesia & S.A. Zenios [Eds.]. *Operations research, models in quantitative finance*. Heidelberg: Physica-Verlag.
- Babbie, E. [1989]. *The practice of social research* [Fifth edition]. Belmont, California: Wadsworth Publishing Company.
- Barrett, G.V., Phillips, J.S. & Alexander, R.A. [1981]. Concurrent and predictive validity designs: a critical reanalysis. *Journal of Applied Psychology*, 66 [1], 1-6.
- Berenson, M.L., Levine, D.M. & Goldstein, M. [1983]. *Intermediate statistical methods and applications; a computer package approach*. Englewood Cliffs, New Jersey: Prentice Hall.
- Binning, J.F. & Barrett, G.V. [1989]. Validity of Personnel decisions: a conceptual analysis of the inferential and evidential bases. *Journal of Applied Psychology*, 74 [3], 478-494.
- Blum, M.L. & Naylor, J.C. [1968]. *Industrial psychology: its theoretical and social foundations*. New York: Harper & Row.
- Bobko, P. [1983]. An analysis of correlations corrected for attenuation and range restriction. *Journal of Applied Psychology*, 68 [4], 584-589.

- Bobko, P. & Rieck, A. [1980]. Large sample estimators for standard errors in functions of correlation coefficients. *Applied Psychological Measurement*, 4 [3], 385-398.
- Boone, J.O. & Lewis, M.A. [1980]. Demonstration of possible effects of recruitment and selection procedures on correcting the validity coefficient for restriction of range. *Psychological Reports*, 46, 927-930.
- Boudreau, J.W. [1983a]. Economic considerations in estimating the utility of human resource productivity improvement programs. *Personnel Psychology*, 36, 551-576.
- Boudreau, J.W. [1983b]. Effects of employee flows on utility analysis of human resource productivity improvement programs. *Journal of Applied Psychology*, 68 [3], 396-406.
- Boudreau, J.W. [1984]. Decision theory contributions to HRM research and practice. *Industrial Relations*, 23 [2], 198-217.
- Boudreau, J.W. [1989]. Selection utility analysis: a review and agenda for further research. In M. Smith & I. Robertson [Eds.]. *Advances in selection and assessment*. Chichester: John Wiley & Sons.
- Boudreau, J.W. [1991]. Utility analysis for decisions in human resource management. In M.D. Dunnette & L.M. Hough [Eds.]. *Handbook of industrial and organizational psychology* [Second edition; Volume 2]. Palo Alto, California: Consulting Psychologists Press, Inc.
- Boudreau, J.W. & Berger C.J. [1985a]. Decision-theoretic utility analysis applied to employee separations and acquisitions. *Journal of Applied Psychology*, 70 [3], 581-612.
- Boudreau, J.W. & Berger, C.J. [1985b]. Towards a model of employee movement utility. In K.M. Rowland & G.R. Ferris [Eds.]. *Research in personnel and human resources management* [Volume 3]. Greenwich, Connecticut: JAI Press Inc.
- Boudreau, J.W. & Rynes, S.L. [1985]. Role of recruitment in staffing utility analysis. *Journal of Applied Psychology*, 70, 354-366.
- Boyatzis, R.E. [1982]. *The competent manager*. New York: Wiley.
- Brogden, H.E. [1946]. On the interpretation of the correlation coefficient as a measure of predictive efficiency. *The Journal of Educational Psychology*, 37, 65-76.
- Brogden, H.E. [1949]. When testing pays off. *Personnel Psychology*, 2, 171-183.
- Campbell, J.P. [1991]. Modeling the performance prediction problem in industrial and organizational psychology. In M.D. Dunnette & L.M. Hough [Eds.]. *Handbook of industrial and organizational psychology* [Second edition; Volume 1]. Palo Alto, California: Consulting Psychologists Press, Inc.
- Campbell, D.T. & Stanley, J.C. [1963]. *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally College Publishing Company.
- Carnap, R. [1950]. *Logical foundation of probability*. Chicago: University of Chicago Press.
- Cascio, W.F. [1991a]. *Applied psychology in personnel management*. Englewood Cliffs, N.J.: Prentice-Hall.
- Cascio, W.F. [1991b]. *Costing human resources; the financial impact of behavior in organizations*. Boston: PWS-Kent Publishing Company.
- Cascio, W.F. & Ramos, R. [1986]. Development and application of a new method for assessing job performance in behavioral/economic terms. *Journal of Applied Psychology*, 1, 20-28.

- Clark, J.J., Hindelang, T.J. & Pritchard, R.E. [1984]. *Capital budgeting: planning and control of capital expenditures*. Englewood Cliffs, New Jersey: Prentice Hall.
- Cleary, T.A. [1968]. Test bias: prediction of grades of Negro and white students in integrated colleges. *Journal of Educational Measurement*, 5, 115-124.
- Cohen, J. [1977]. *Statistical power analysis for the behavioral sciences* [Revised edition]. New York: Academic Press.
- Cohen, J. & Cohen, P. [1983]. *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, New Jersey: Lawrence Erlbaum Associates Publishers.
- Cook, T.D. & Campbell, D.T. [1979]. *Quasi-experimentation: design and analysis issues for field settings*. Chicago: Rand McNally.
- Cook, T.D., Campbell, D.T. & Peracchio, L. [1991]. Quasi experimentation. In M.D. Dunnette & L.M. Hough [Eds.]. *Handbook of industrial and organizational psychology* [Second edition; Volume 2]. Palo Alto, California: Consulting Psychologists Press, Inc.
- Coombs, [1950]. *A theory of data*. New York: Wiley and Sons.
- Copi, I.M. & Cohen, C. [1990]. *Introduction to logic*. New York: Macmillan Publishing Company.
- Coxon, A.P.M. [1982]. *The user's guide to multidimensional scaling*. London: Heinemann Educational Books.
- Crocker, L. & Algina, J. [1986]. *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.
- Cronbach, L. [1951]. Coefficient alpha and the internal structure of tests. *Psychometrika*, 16 [3], 297-333.
- Cronbach, L.J. & Gleser, G.C. [1965]. *Psychological tests and personnel decisions* [2nd edition]. Urbana, Illinois: University of Illinois Press.
- Cronbach, L.J., Gleser, G.C., Nanda, H. & Rajaratnam, N. [1972]. *The dependability of behavioral measurements*. New York: John Wiley.
- Cronshaw, S.F. & Alexander, R.A. [1985]. One answer to the demand for accountability: selection utility as an investment decision. *Organizational Behavior and Human Decision Processes*, 35, 102-118.
- Crous, M.J. [1986]. Personeelbestuur. In P.G. Du Plessis [Red.]. *Toegepaste bedryfseconomie: 'n inleidende oorsig*. Pretoria: HAUM Uitgewery.
- Diaconis, P.M. & Efron, B.S. [1983]. Computer-intensive methods in statistics. *Scientific American*, 116-130.
- Dobson, P. [1988]. The correction of correlation coefficients for restriction of range when restrictions result from the truncation of a normally distributed variable. *British Journal of Mathematical and Statistical Psychology*, 41, 227-234.
- Drasgow, F. & Hulin, C.L. [1991]. Item response theory. In M.D. Dunnette & L.M. Hough [Eds.]. *Handbook of industrial and organizational psychology* [Second edition; Volume 2]. Palo Alto, California: Consulting Psychologists Press, Inc.
- Duvenage, A. [1991]. *A heuristic on the validity generalization dilemma*. HSRC Report R/Pers 786. Pretoria: Human Sciences Research Council.
- Efron, B.S. [1982]. *The jackknife, the bootstrap and other resampling plans*. Philadelphia: Society for Industrial and Applied Mathematics.
- Efron, B.S. [1988]. Bootstrap confidence intervals: good or bad? *Psychological Bulletin*, 104, 293-296.



- Efron, B.S. & Gong, G. [1983]. A leisurely look at the bootstrap, the jackknife and cross-validation. *The American Statistician*, 37 [1], 36-48.
- Efron, B.S. & Tibshirani, R.J. [1986]. Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy. *Statistical Science*, 1, 54-74.
- Efron, B.S. & Tibshirani, R.J. [1993]. *An introduction to the bootstrap*. New York: Chapman and Hall.
- Einhorn, H.J. & Bass, A.R. [1971]. Methodological considerations relevant to discrimination in employment testing. *Psychological Bulletin*, 75 [4], 261-269.
- Ellis, M.V. & Blustein, D.L. [1991]. Developing and using educational and psychological tests and measures: the unificationist perspective. *Journal of Counseling and Development*, 69, 550-555.
- Feldt, L.S. [1975]. Estimation of the reliability of a test divided into two parts of unequal length. *Psychometrika*, 40 [4], 557-561.
- Feldt, L.S. & Brennan, R.L. [1989]. Reliability. In R.L. Linn [Ed.]. *Educational Measurement* [Third Ed.]. New York: American Council on Education.
- Fitz-Enz, J. [1980]. Quantifying the human resource function. *Personnel*, 57 [3], 41-52.
- Florin-Thuma, B.C. & Boudreau, J.W. [1987]. Performance feedback utility in a small organization: effects on organizational outcomes and managerial decision processes. *Personnel Psychology*, 40, 693-713.
- Forsythe, R.A. [1971]. An empirical note on a correlation coefficient corrected for restriction of range. *Educational and Psychological Measurement*, 31, 115-123.
- Forsyth, A. & Feldt, L.S. [1969]. An investigation of empirical sampling distributions of correlation coefficients corrected for attenuation. *Educational and Psychological Measurement*, 29, 61-71.
- Gatewood, R.B. & Feild, H.S. [1994]. *Human resource selection*. Fort Worth: The Dryden Press.
- Ghiselli, E.E., Campbell, J.P. & Zedeck, S. [1981]. *Measurement theory for the behavioural sciences*. San Francisco: W.H. Freeman and Company.
- Gilmer, J.S. & Feldt, L.S. [1983]. Reliability estimation for a test with parts of unknown lengths. *Psychometrika*, 48 [1], 99-111.
- Gow, J.F. [1985]. Human Resource managers must remember the bottom line. *Personnel Journal*, 30-32.
- Greener, J.M. & Osburn, H.G. [1980]. Accuracy of corrections for restriction in range due to explicit selection in heteroscedastic and nonlinear distributions. *Educational and Psychological Measurement*, 40, 337-346.
- Gross, A.L. [1982]. Relaxing the assumptions underlying corrections for restriction of range. *Educational and Psychological Measurement*, 42, 795-801.
- Gross, A.L. [1990]. A maximum likelihood approach to test validation with missing and censored dependent variables. *Psychometrika*, 55 [3], 533-549.
- Gross, A.L. & Fleishman, L. [1983]. Restriction of range corrections when both distribution and selection assumptions are violated. *Applied Psychological Measurement*, 7 [2], 227-237.
- Gross, A.L. & Kagen, E. [1983]. Not correcting for restriction of range can be advantageous. *Educational and Psychological Measurement*, 43, 389-396.
- Gross, A.L. & McGanney, M.L. [1987]. The restriction of range problem and non-ignorable selection processes. *Journal of Applied Psychology*, 72 [4], 604-610.



- Gross, A.L. & Su, W. [1975]. Defining a fair or unbiased selection model: a question of utilities. *Journal of Applied Psychology*, 60 [3], 345-351.
- Guilford, J.P. & Fruchter, B. [1978]. *Fundamental statistics in psychology and education*. Tokyo: McGraw-Hill.
- Guion, R.M. [1965]. *Personnel testing*. New York: McGraw-Hill Book Company.
- Guion, R.M. [1966]. Employment tests and discriminatory hiring. *Industrial Relations*, 5, 20-37.
- Guion, R.M. [1991]. Personnel assessment, selection and placement. In M.D. Dunnette & L.M. Hough [Eds.]. *Handbook of industrial and organizational psychology* [Second edition; Volume 2]. Palo Alto, California: Consulting Psychologists Press, Inc.
- Guion, R.M. & Cranny, C.J. [1982]. A note on concurrent and predictive validity designs: a critical reanalysis. *Journal of Applied Psychology*, 67 [2], 239-244.
- Gulliksen, H. [1950]. *Theory of mental tests*. New York: John Wiley and Sons.
- Gullickson, A & Hopkins, K. [1976]. Interval estimation of correlation coefficients corrected for restriction of range. *Educational and Psychological Measurement*, 36, 9-25.
- Guttman, L. [1945]. A basis for analyzing test-retest reliability. *Psychometrika*, 10 [4], 255-282.
- Hakstian, A.R., Schroeder, M.L. and Rogers, W.T. [1988]. Inferential procedures for correlation coefficients corrected for attenuation. *Psychometrika*, 53 [1], 27-43.
- Hakstian, A.R., Schroeder, M.L. and Rogers, W.T. [1989]. Inferential theory for partially disattenuated correlation coefficients. *Psychometrika*, 54 [3], 397-407.
- Hambleton, R.K. [1989]. Principles and selected applications of item response theory. In R.L. Linn [Ed.]. *Educational Measurement* [Third Ed.]. New York: American Council on Education.
- Hanges, P.I., Rentsch, J.R., Yusko, K.P. & Alexander, R.A. [1991]. Determining the appropriate correction when the type of range restriction is unknown: developing a sample-based procedure. *Educational and Psychological Measurement*, 51, 329-340.
- Hays, W.L. [1971]. *Statistics for the social sciences* [second edition]. London: Holt, Rinehart and Winston.
- Held, J.D. & Foley, P.P. [1994]. Explanations for accuracy of the general multivariate formulas in correcting for range restriction. *Applied Psychological Measurement*, 18 [4], 355-367.
- Holmes, D.J. [1990]. The robustness of the usual corrections for restriction in range due to explicit selection. *Psychometrika*, 55 [1], 19-32.
- Horst, P. [1951]. Estimating the total test reliability from parts of unequal length. *Educational and Psychological Measurement*, 11, 368-371.
- Hoyt, C. [1941]. Test reliability estimated by analysis of variance. *Psychometrika*, 6 [3], 1941.
- Huber, G.P. [1980]. *Managerial decision making*. Glenview, Illinois: Scott, Foresman.
- Huck, S.W. [1992]. Group heterogeneity and Pearson's r. *Educational and Psychological Measurement*, 52, 253-260.
- Hull, C.L. [1928]. *Aptitude testing*. Yonkers, NY: World Book.
- Hunter, J.E. & Schmidt, F.L. [1976]. Critical analysis of the statistical and ethical implications of various definitions of test bias. *Psychological Bulletin*, 83 [6], 1053-1071.
- Hunter, J.E. & Schmidt, F.L. [1982]. Fitting people to jobs: the impact of personnel selection on national productivity. In M.D. Dunnette & E.A. Fleischman [Eds.]. *Human performance and productivity* [Vol. 1, pp.233-284]. Hillsdale, New Jersey: Erlbaum.

- Hunter, J.E., Schmidt, F.L. & Jackson, G.B. [1982]. *Meta-analysis: cumulating research findings across studies*. Beverly Hills, CA.: Sage.
- Johnston, J.T. & Ree, M.J. [1994]. Rangej: a Pascal program to compute the multivariate correction for range restriction. *Educational and Psychological Measurement*, 54 [3], 693-695.
- Kane, M.T. [1992]. An argument-based approach to validity. *Psychological Bulletin*, 112 [3], 527-535.
- Keats, J.A. [1957]. Estimation of error variances of test scores. *Psychometrika*, 22, 29-41.
- Keeney, R.L. & Raiffa, H. [1976]. *Decisions with multiple objectives: preferences and value tradeoffs*. New York: John Wiley & Sons.
- Kelley, T.L. [1923]. *Statistical method*. New York: The MacMillan Company.
- Kelley, T.L. [1947]. *Fundamentals of statistics*. Cambridge: Harvard University Press.
- Kendall, M.G. & Stuart, A. [1977]. *The advanced theory of statistics* [Vol. 1, Fourth ed.] London: Charles Griffin and Company.
- Kerlinger, F.N. [1986]. *Foundations of behavioural research*. New York: CBS Publishing.
- Kleiman, L.S. & Faley, R.H. [1985]. The implications of professional and legal guidelines for court decisions involving criterion related validity: a review and analysis. *Personnel Psychology*, 38, 803-833.
- Kleinbaum, D.G. & Kupper, L.L. [1978]. *Applied regression analysis and other multivariate methods*. North Scituate, Massachusetts: Duxbury Press.
- Kristof, W. [1974]. Estimation of reliability and true score variance from a split of a test into three arbitrary parts. *Psychometrika*, 39 [4], 491-499.
- Kristof, W. [1982]. *Contributions to the analysis of correlation coefficients*. Unpublished doctoral dissertation, University of South Africa, Pretoria.
- Kuder, F. [1991]. Comments concerning the appropriate use of formulas for estimating the internal consistency reliability of tests. *Educational and Psychological Measurement*, 51, 873-874.
- Kuder, G.F. & Richardson, M.W. [1937]. The theory of the estimation of test reliability. *Psychometrika*, 2 [3], 151-160.
- Landy, F.J. [1986]. Stamp collecting versus science; validation as hypothesis testing. *American Psychologist*, 41 [11], 1183-1192.
- Landy, F.J., Farr, J.L. & Jacobs, R.R. [1982]. Utility concepts in performance measurement. *Organizational Behavior and Human Performance*, 30, 15-40.
- Lawsche, C.H. & Balma, M.J. [1966]. *Principles of personnel testing*. New York: McGraw-Hill Book Company.
- Lee, R. & Foley, P. [1986]. Is the validity of a test constant throughout the test score range?. *Journal of Applied Psychology*, 71 [4], 641-644.
- Lee, R., Miller, K.J. & Graham, W.K. [1982]. Corrections for restriction of range and attenuation in criterion-related validation studies. *Journal of Applied Psychology*, 67 [5], 637-639.
- Levin, J. [1972]. The occurrence of an increase in correlation by restriction of range. *Psychometrika*, 37 [1], 93-97.
- Levy, H. & Sarnat, M. [1994]. *Capital investment and financial decisions*. New York: Prentice Hall.

- Linn, R.L. [1968]. Range restriction problems in the use of self-selected groups for test validation. *Psychological Bulletin*, 69 [1], 69-73.
- Linn, R.L. [1983]. Pearson selection formulas: implications for studies of predictive bias and estimates of educational effects in selected samples. *Journal of Educational Measurement*, 20 [1], 1-15.
- Linn, R.L., Harnish, D.L. & Dunbar, S.B. [1981]. Corrections for range restriction: an empirical investigation of conditions resulting in conservative corrections. *Journal of Applied Psychology*, 66 [6], 655-663.
- Lipsey, M.W. [1990]. *Design sensitivity; statistical power for experimental research*. Newbury Park, California: Sage Publications.
- Lord, F.M. & Novick, M.R. [1968]. *Statistical theories of mental test scores*. Reading, Massachusetts: Addison-Wesley Publishing Company.
- Lumby, S. [1994]. *Investment appraisal and financial decisions* [Fifth edition]. London: Chapman & Hall.
- Lunneborg, C.E. [1985]. Estimating the correlation coefficient: the bootstrap approach. *Psychological Bulletin*, 98, 209-215.
- March, J.G. & Simon, H.G. [1958]. *Organizations*. New York; Wily.
- Margenau, H. [1950]. *The nature of physical reality*. New York: McGraw-Hill.
- Marx, F.W. [1983]. Personeelbestuur. In F.W. Marx & P.J. Van Aswegen [Reds.]. *Die bedryfseksonomie: 'n kort oorsig*. Pretoria: HAUM Uitgewery.
- Maxwell, S.E. & Arvey, R.D. [1993]. The search for predictors with high validity and low adverse impact: compatible or incompatible goals?. *Journal of Applied Psychology*, 78 [3], 433-437.
- Mayer, J.D. Estimating the sampling variance of correlation corrected for attenuation using coefficient alpha. *Perceptual and Motor Skills*, 56, 207-210.
- McGuire, D.P. [1986]. The occurrence of an increase in correlation with explicit selection. *Psychometrika*, 51 [2], 331-333.
- McGuire, W.J. [1997]. Creative hypothesis generating in psychology: some useful heuristics. *Annual Review of Psychology*, 48, 1-30.
- Mendoza, J.L. [1993]. Fisher transformations for correlations corrected for selection and missing data. *Psychometrika*, 58 [4], 601-615.
- Mendoza, J.L. & Mumford, M. [1987]. Corrections for attenuation and range restriction on the predictor. *Journal of Educational Statistics*, 12 [3], 282-293.
- Mendoza, J.L., Hart D.E. & Powell, A. [1991]. A bootstrap confidence interval based on a correlation corrected for range restriction. *Multivariate Behavioral Research*, 26 [2], 255-269.
- Messick, S. [1989]. *Validity*. In R.L. Linn [Ed.]. *Educational measurement* [Third edition]. New York: American Council on Education and McMillan Publishing Company.
- Milkovich, G.T. & Boudreau, J.W. [1994]. *Human resource management* [Seventh edition]. Homewood, Illinois: Richard D. Irwin Inc.
- Moran, P.A.P. [1970]. On asymptotically optimal tests of composite hypothesis. *Biometrika*, 57 [1], 47-55.
- Mouton, J. & Marais, H.C. [1985]. *Metodologie van die geesteswetenskappe: basiese begrippe*. Pretoria: Raad vir Geesteswetenskaplike Navorsing.

- Muchinsky, P.M. [1983]. *Psychology applied to work; an introduction to industrial and organizational psychology*. Homewood, Illinois; The Dorsey Press.
- Naylor, J.C. & Shine, L.C. [1965]. A table for determining the increase in mean criterion score obtained by using a selection device. *Journal of Industrial Psychology*, 3, 33-42.
- Novick, M.R. [1966]. The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, 3, 1-18.
- Novick, M.R. & Lewis, C. [1967]. Coefficient alpha and the reliability of composite measurements. *Psychometrika*, 32 [1], 1-13.
- Nunnally, J.C. [1978]. *Psychometric theory*. New York: McGraw-Hill Book Company.
- Olson, C.A. & Becker, B.E. [1983]. A proposed technique for the treatment of restriction of range in selection validation. *Psychological Bulletin*, 93 [1], 137-148.
- Owen, K. & Taljaard, J.J. [1988]. Pretoria: Human Sciences Research Council
- Pearson K. [1903]. Mathematical contributions to the theory of evolution XI. On the influence of natural selection on the variability and correlation of organs. *Philosophical Transactions*, 200, 1-66.
- Petersen N.S. [1976]. An expected utility model for optional selection. *Journal of Educational Statistics*, 1 [4], 333-358.
- Popper, K.R. [1972]. *Conjectures and refutations; the growth of scientific knowledge*. London: Routledge and Paul.
- Rademeyer, W.F. [1983]. Finansiële bestuur. In F.W. Marx & P.J. Van Aswegen [Reds.]. *Die bedryfsekonomie: 'n kort oorsig*. Pretoria: HAUM Uitgewery.
- Raju, N.S. [1970]. New formula for estimating total test reliability from parts of unequal lengths. *Proceedings of the 78th Annual Convention of the American Psychological Association*, 5, 143-144.
- Raju, N.S. [1977]. A generalization of coefficient alpha. *Psychometrika*, 42 [4], 549-565.
- Raju, N., Burke, M.J. & Maurer, T.J. [1995]. A note on direct range restriction corrections in utility analysis. *Personnel Psychology*, 48, 143-149.
- Raju, N.S., Edwards, J.E. & LoVerde, M.A. [1985]. Corrected formulas for computing sample sizes under indirect range restriction. *Journal of Applied Psychology*, 70 [3], 565-566.
- Raju, N.S., Normand, J. & Burke, M.J. [1990]. A new approach for utility analysis. *Journal of Applied Psychology*, 75 [1], 3-12.
- Rao, C.R. [1973]. *Linear statistical inference and its application*. New York: John Wiley & Sons.
- Rasmussen, J.L. [1987]. Estimating correlation coefficients: bootstrap and parametric approaches. *Psychological Bulletin*, 101, 136-139.
- Rasmussen, J.L. [1988]. "Bootstrap confidence intervals: good or bad": comments on Efron [1988] and Strube [1988] and further evaluation. *Psychological Bulletin*, 104, 297-299.
- Raudenbush, S.W. & Bryk, A.S. [1985]. Empirical Bayes meta-analysis. *Journal of Educational Statistics*, 10, 75-98.
- Rauschenberger, J.M. & Schmidt, F.L. [1987]. Measuring the economic impact of human resource programs. *Journal of Business and Psychology*, 2, 50-59.



- Ree, M.J., Carretta, T.R., Earles, J.A. & Albert, W. [1994]. Sign changes when correcting for range restriction: a note on Pearson's and Lawley's selection formulas. *Journal of Applied Psychology*, 79 [2], 289-301.
- Rich, J.R. & Boudreau, J.W. [1987]. The effects of variability and risk in selection utility analysis: an empirical comparison. *Personnel Psychology*, 40, 55-84.
- Rogers, W.T. [1976]. Jackknifing disattenuated correlations. *Psychometrika*, 41, 121-133.
- Rossi, P.H. & Freeman, H.E. [1985]. *Evaluation, a systematic approach* [Third edition]. Beverly Hills, California: Sage Publications.
- Rulon, P.J. [1939]. A simplified procedure for determining the reliability of a test by split-halves. *Harvard Educational Review*, 9, 99-103.
- Sackett, P.R. & Wade, B.E. [1983]. On the feasibility of criterion-related validity: the effects of range restriction assumptions on needed sample size. *Journal of Applied Psychology*, 68 [3], 374-381.
- Sands, W.A., Alf, E.F. & Abrahams, N.M. [1978]. Correction of validity coefficients for direct restriction in range occasioned by univariate selection. *Journal of Applied Psychology*, 63 [6], 747-750.
- SAS Institute [1988]. *SAS/STAT user's guide* [Release 6.03 edition]. Cary, NC.: SAS Institute Inc.
- Saville and Holdsworth Ltd. [1996]. *Competency modelling*. Unpublished manuscript, SHL South Africa, Groenkloof.
- Schepers, J.M. [1996]. The development of a statistical procedure to correct the effects of restriction of range on validity coefficients. *Journal of Industrial Psychology*, 22 [1], 19-27.
- Schiffman, S.S., Reynolds, M.L. & Young F.W. [1981]. *Introduction to multidimensional scaling: theory, methods and applications*. New York: Academia Press.
- Schmidt, F.L. & Hunter, J.E. [1977]. Development of a general solution to the problem of validity generalisation. *Journal of Applied Psychology*, 62 [5], 529-540.
- Schmidt, F.L., Hunter, J.E. & Urry, V.W. [1976]. Statistical power in criterion-related validation studies. *Journal of Applied Psychology*, 61 [4], 473-485.
- Schmidt, F.L., Hunter, J.E., McKenzie, R.C. & Muldrow, T.W. [1979]. Impact of valid selection procedures on work-force productivity. *Journal of Applied Psychology*, 64, 609-626.
- Schmidt, F.L., Hunter, J.E., Pearlman, K. & Hirsh, H.R. [1985]. Forty questions about validity generalization and meta-analysis. *Personnel Psychology*, 38, 697-798.
- Schmitt, N. & Landy, F.J. [1993]. The concept of validity. In N. Schmitt, W. Borman and Associates [Eds.]. *Personnel selection in organizations*. San Francisco: Jossey-Bass Publishers.
- Schmitt, N. & Landy, F.J. The concept of validity. In N. Schmitt, W.C. Borman & Associates. *Personnel selection in organizations*. San Francisco: Jossey-Bass Publishers.
- Schuler, H. & Guldin, A. [1991]. Methodological issues in personnel selection research. In C.L. Cooper & I.T. Robertson [Eds.]. *International Review of Industrial and Organizational Psychology*. New York: John Wiley & Sons
- Shao, J. & Tu, D. [1995]. *The jackknife and bootstrap*. New York: Springer-Verlag.
- Shavelson, R.J. & Webb, N.M. [1991]. *Generalizability theory; a primer*. London; Sage Publications.

- Sheppeck, M.A. & Cohen, S.L. [1985]. Put a dollar value on your training programs. *Training and Development Journal*, 39, 59-62.
- Singer, M. [1993]. *Fairness in personnel selection*. Avebury: Aldershot.
- Society for Industrial and Organizational Psychology. [1987]. *Principles for the validation and use of personnel selection procedures* [Third edition]. College Park, MD: Author.
- Society for Industrial Psychology. [1992]. *Guidelines for the validation and use of personnel selection procedures*. Lynwoodrif: Author.
- Spangenberg, H. [1990]. *Assessing managerial competence*. Kenwyn: Juta & Company.
- Spearman, C. [1904]. The proof and measurement of association between two things. *American Journal of Psychology*, 15, 72-101
- Spearman, C [1907]. Demonstration of formulae for true measurement of correlation. *American Journal of Psychology*, 18, 161-169.
- Spearman, C. [1910]. Correlation calculated from faulty data. *British Journal of Psychology*, 3, 271-295.
- Spearman, C. [1913]. Correlation of sums and differences. *British Journal of Psychology*, 5, 417-426.
- Stanley, J.C. [1971]. Reliability. In R.L. Thorndike [Ed.]. *Educational Measurement* [Second Ed.]. Washington, DC: American Council on Education.
- Strube, M.J. [1988]. Bootstrap type I error rates for the correlation coefficient: an examination of alternative procedures. *Psychological Bulletin*, 104, 290-292.
- Sussmann, M. & Robertson, D.U. [1986]. The validity of validity: an analysis of validation study designs. *Journal of Applied Psychology*, 71 [3], 461-468.
- Tabachnick, B.G. & Fidell, L.S. [1989]. *Using multivariate statistics* [Second edition]. New York: Harper Collins Publishers.
- Taylor, H.C. & Russell, J.T. [1939]. The relationship of validity coefficients to the practical effectiveness of tests in selection: discussion and tables. *Journal of Applied Psychology*, 23, 565-578.
- Ten Berge, J.M.F. & Zegers, F.E. [1978]. A series of lower bounds to the reliability of a test. *Psychometrika*, 43 [4], 575-579.
- Tenopyr, M.L. [1983]. The realities of employment testing. In K.M. Rowland, G.R. Ferris & J.L. Sherman [Eds]. *Current issues in personnel management* [Second edition]. Boston: Allyn and Bacon, Inc.
- Thompson, P.A. [1991]. Resampling approaches to complex psychological experiments. *Multivariate Behavioral Research*, 26 [4], 737-763.
- Thorndike, R.L. [1949]. *Personnel selection; test and measurement techniques*. New York: John Wiley and Sons.
- Thorndike, R.L. [1971]. Concepts of culture fairness. *Journal of Educational Measurement*, 8, 63-70.
- Thorndike, R.L. [1982]. *Applied Psychometrics*. Boston: Houghton Mifflin Company.
- Tiffen, J. [1946]. *Industrial Psychology*. Englewood Cliffs, N.J.: Prentice Hall.
- Toothaker, L.E. [1986]. *Introductory Statistics for the behavioral sciences*. New York: McGraw-Hill Book Company
- Torgerson, W.S. [1958]. *Theory and methods of scaling*. New York: John Wiley and Sons.
- Tucker, L.R. [1949]. A note on the estimation of test reliability by the Kuder-Richardson formula-20. *Psychometrika*, 14, 117-119.





**APPENDIX A**  
**SAS PROGRAM LISTINGS**

A1. THE EFFECT OF PARTIALLY CORRECTING THE CORRELATION COEFFICIENT FOR ATTENUATION

```
OPTIONS LINESIZE=80 NODATE NONUMBER;
DATA DAT1;
DO RXY=.05 TO 1 BY 0.05;
DO RTTY=.05 TO 1 BY 0.05;
OUTPUT;
END;
END;
DATA DAT2;
SET DAT1;
IF RXY GT SQRT(RTTY) THEN RTTY=.;
RcXY=RXY/SQRT(RTTY);
DIFF=RCXY-RXY;
FILENAME XYZ 'PIC1.GRAF';
GOPTIONS
DEVICE=HPLJGL
AUTOFEED
GSFMODE=REPLACE
HANDSHAKE=XONXOFF
NOPROMPT
NODISPLAY
NOCELL
NOCHARACTERS
GSFNAME=XYZ;
PROC G3D;
SCATTER RXY*RTTY=RcXY/ROTATE=10 30 50;
RUN;
```

**A2. THE EFFECT OF CORRECTING THE CORRELATION COEFFICIENT FOR CASE B  
RESTRICTION OF RANGE**

```
OPTIONS LINESIZE=80 NODATE NONUMBER;
DATA DAT1;
DO RXY=.03 TO 1 BY 0.03;
DO K=1 TO 4 BY 0.2;
OUTPUT;
END;
END;
DATA DAT2;
SET DAT1;
RcXY=(RXY*K)/(SQRT(1-RXY**2)+(RXY**2)*(K**2));
DIFF=RCXY-RXY;
FILENAME XYZ 'PIC2.GRAF';
GOPTIONS
DEVICE=HPLJGL
AUTOFEED
GSFMODE=REPLACE
HANDSHAKE=XONXOFF
NOPROMPT
NODISPLAY
NOCELL
NOCHARACTERS
GSFNAME=XYZ;
PROC G3D;
SCATTER RXY*K=RcXY;
RUN;
```

### A3. THE EFFECT OF CORRECTING THE CORRELATION COEFFICIENT FOR CASE C RESTRICTION OF RANGE

```

OPTIONS LINESIZE=80 NODATE NONUMBER;
DATA DAT1;
DO RXY=0 TO 1 BY 0.04;
DO RXZ=.5, .8;
DO K=1 TO 5 BY 0.2;
RYZ = RXZ;
OUTPUT;
END;
END;
END;
DATA DAT2;
SET DAT1;
DD1=((RXZ**2)+(RYZ**2))-(RXY**2);
DD2=((RXY**2)+(RYZ**2))-(RXZ**2);
DD3=((RXY**2)+(RXZ**2))-(RYZ**2);
IF DD1 GT 1 OR DD2 GT 1 OR DD3 GT 1 THEN RXY = .;
RcXY=((RXY-RXZ*RYZ)+(RXZ*RYZ)*(K**2))/SQRT(((1-(RXZ**2)+(RXZ**2)*(K**2))*(1-
(RYZ**2)+(RYZ**2)*(K**2))));
DIFF=RCXY-RXY;
PROC SORT;
BY RXZ RYZ;
FILENAME XYZ 'PIC3.GRAF';
GOPTIONS
DEVICE=HPLJGL
AUTOFEED
GSFMODE=REPLACE
HANDSHAKE=XONXOFF
NOPROMPT
NODISPLAY
NOCELL
NOCHARACTERS
GSFNAME=XYZ;
PROC G3D;
SCATTER RXY*K=RcXY;
BY RXZ RYZ;
RUN;

```

A4. THE EFFECT OF CORRECTING THE CORRELATION COEFFICIENT FOR CASE A  
RESTRICTION OF RANGE

```
OPTIONS LINESIZE=80 NODATE NONUMBER;
DATA DAT1;
DO RXY=.01 TO 1 BY 0.02;
DO K=.1 TO 1 BY .1;
OUTPUT;
END;
END;
DATA DAT2;
SET DAT1;
RcXY=SQRT(1-((K**2)*(1-(RXY**2))));
DIFF=RCXY-RXY;
FILENAME XYZ 'PIC4.GRAF';
GOPTIONS
DEVICE=HPLJGL
AUTOFEED
GSFMODE=REPLACE
HANDSHAKE=XONXOFF
NOPROMPT
NODISPLAY
NOCELL
NOCHARACTERS
GSFNAME=XYZ;
PROC G3D;
SCATTER RXY*K=RcXY/ROTATE=10;
RUN;
```



**A5. THE EFFECT OF DOUBLE CORRECTING THE CORRELATION COEFFICIENT FOR  
CRITERION UNRELIABILITY AND CASE B RESTRICTION OF RANGE**

```

OPTIONS LINESIZE=80 NODATE NONUMBER;
DATA DAT1;
DO RXY=.05 TO 1 BY 0.05;
DO RTTY=.1 TO .9 BY .05;
DO K=4, 5;
OUTPUT;
END;
END;
END;
DATA DAT2;
SET DAT1;
IF RXY GT SQRT(RTTY) THEN RXY=.;
RcXY=(RXY*(RTTY**-.5)*K)/(SQRT(1-((RXY**2)/RTTY)+(((RXY**2)/RTTY)*(K**2))));
DIFF=RCXY-RXY;
PROC SORT;
BY K;
FILENAME XYZ 'PIC5.GRAF';
GOPTIONS
DEVICE=HPLJGL
AUTOFEED
GSFMODE=REPLACE
HANDSHAKE=XONXOFF
NOPROMPT
NODISPLAY
NOCELL
NOCHARACTERS
GSFNAME=XYZ;
PROC G3D;
SCATTER RXY*RTTY=RcXY;
BY K;
RUN;

```

**A6. THE STANDARD ERROR OF THE CORRELATION COEFFICIENT CORRECTED FOR  
CRITERION UNRELIABILITY [KELLEY FORMULA]**

```

OPTIONS LINESIZE=80 NODATE NONUMBER;
DATA DAT1;
DO RXY=.04 TO 1 BY 0.04;
DO RTTYS=.1 TO .9 BY .05;
DO n=10, 90;
OUTPUT;
END;
END;
END;
DATA DAT2;
SET DAT1;
RTTY=(2*RTTYS)/(1+RTTYS);
IF RXY GT SQRT(RTTY) THEN RXY=.;
RcXY=RXY/(SQRT(RTTY));
SEr=(1-(RXY**2))/SQRT(n);
SErc=SQRT(((RcXY**2)/(n-2))*((RcXY**2)+(1/(RXY**2))-(1/RTTYS)+(1/(4*(RTTYS**2)))-(5/4)));
Z1=RXY/SEr;
Z2=RcXY/SErc;
G=Z2/Z1;
V=SErc/SEr;
Zk=1.6449;
IF Z1 GE Zk AND Z2 GE Zk THEN GG=1;
IF Z1 GE Zk AND Z2 LT Zk THEN GG=2;
IF Z1 LT Zk AND Z2 GE Zk THEN GG=3;
IF Z1 LT Zk AND Z2 LT Zk THEN GG=4;
IF Z1=. OR Z2=. THEN GG=.;
IF GG=1 THEN SHAPEVAR='CUBE';
IF GG=2 THEN SHAPEVAR='HEART';
IF GG=3 THEN SHAPEVAR='SPADE';
IF GG=4 THEN SHAPEVAR='FLAG';
PROC FORMAT;
VALUE GGF 1='VH01VH02'
        2='VH01AH02'
        3='AHO1VH02'
        4='AH01AH02';
FORMAT GG GGF.;
PROC SORT;
BY n;
FILENAME XYZ 'PIC6.GRAF';

```

```
GOPTIONS
DEVICE=HPLJGL
AUTOFEED
GSFMODE=REPLACE
HANDSHAKE=XONXOFF
NOPROMPT
NODISPLAY
NOCELL
NOCHARACTERS
GSFNAME=XYZ;
PROC G3D;
SCATTER RT*TY* RXY=SErc/ROTATE=10;
BY n;
RUN;
```

**A7. THE REACTION OF G TO CHANGES IN THE VALIDITY COEFFICIENT, THE CRITERION RELIABILITY COEFFICIENT AND SAMPLE SIZE, USING BOBKO & RIECK'S THIRD FORMULA FOR THE STANDARD ERROR OF THE CORRECTED CORRELATION COEFFICIENT**

```

OPTIONS LINESIZE=80 NODATE NONUMBER;
DATA DAT1;
DO RXY=.04 TO 1 BY 0.04;
DO RTTY=.1 TO .9 BY .05;
DO n=10, 90;
OUTPUT;
END;
END;
END;
DATA DAT2;
SET DAT1;
IF RXY GT SQRT(RTTY) THEN RXY=.;
RcXY=RXY/SQRT(RTTY);
SEr=(1-(RXY**2))/SQRT(n);
SErc=SQRT((1/n)*(1/RTTY)*((1-RXY**2)**2));
Z1=RXY/SEr;
Z2=RcXY/SErc;
G=Z2/Z1;
V=SErc/SEr;
Zk=1.6449;
IF Z1 GE Zk AND Z2 GE Zk THEN GG=1;
IF Z1 GE Zk AND Z2 LT Zk THEN GG=2;
IF Z1 LT Zk AND Z2 GE Zk THEN GG=3;
IF Z1 LT Zk AND Z2 LT Zk THEN GG=4;
IF Z1=. OR Z2=. THEN GG=.;
IF GG=1 THEN SHAPEVAR='CUBE';
IF GG=2 THEN SHAPEVAR='HEART';
IF GG=3 THEN SHAPEVAR='SPADE';
IF GG=4 THEN SHAPEVAR='FLAG';
PROC FORMAT;
VALUE GGF 1='VH01VH02'
          2='VH01AH02'
          3='AHO1VH02'
          4='AHO1AH02';
PROC SORT;

```

```
BY n;
FILENAME XYZ 'PIC7.GRAF';
GOPTIONS
DEVICE=HPLJGL
AUTOFEED
GSFMODE=REPLACE
HANDSHAKE=XONXOFF
NOPROMPT
NODISPLAY
NOCELL
NOCHARACTERS
GSFNAME=XYZ;
PROC G3D;
SCATTER RTTY*RXY=SErc;
BY n;
DATA DAT3;
SET DAT2;
IF RTTY EQ .1;
PROC PRINT;
RUN;
```

**A8. THE REACTION OF G TO CHANGES IN THE VALIDITY COEFFICIENT, THE CRITERION RELIABILITY COEFFICIENT AND SAMPLE SIZE, USING BOBKO & RIECK'S FIRST FORMULA FOR THE STANDARD ERROR OF THE CORRECTED CORRELATION COEFFICIENT**

```

OPTIONS LINESIZE=80 NODATE NONUMBER;
DATA DAT1;
DO RXY=.04 TO 1 BY 0.04;
DO RTTY=.1 TO .9 BY .05;
DO n=10, 90;
OUTPUT;
END;
END;
END;
DATA DAT2;
SET DAT1;
IF RXY GT SQRT(RTTY) THEN RXY=.;
RcXY=RXY/SQRT(RTTY);
SEr=(1-(RXY**2))/SQRT(n);
A=SQRT((1/n)*(1/RTTY)*((1-RXY**2)**2));
B=(.25/n)*(1/(RTTY**3))*(RXY**2)*((1-(RTTY**2))**2);
C=((1/n)*(1/(RTTY**2))*(RXY)*((RTTY)*(1-(2*(RXY**2)))-(.5*(RXY**2)*(1-(2*(RXY**2))-(RTTY**2)))));
SErc=SQRT(A+B-C);
Z1=RXY/SEr;
Z2=RcXY/SErc;
G=Z2/Z1;
V=SErc/SEr;
Zk=1.6449;
IF Z1 GE Zk AND Z2 GE Zk THEN GG=1;
IF Z1 GE Zk AND Z2 LT Zk THEN GG=2;
IF Z1 LT Zk AND Z2 GE Zk THEN GG=3;
IF Z1 LT Zk AND Z2 LT Zk THEN GG=4;
IF Z1=. OR Z2=. THEN GG=.;
IF GG=1 THEN SHAPEVAR='CUBE';
IF GG=2 THEN SHAPEVAR='HEART';
IF GG=3 THEN SHAPEVAR='SPADE';
IF GG=4 THEN SHAPEVAR='FLAG';
PROC FORMAT;
VALUE GGF 1='VH01VH02'
      2='VH01AH02'
      3='AHO1VH02'
      4='AH01AH02';

```



```
PROC SORT;
BY n;
FILENAME XYZ 'PIC8.GRAF';
GOPTIONS
DEVICE=HPLJGL
AUTOFEED
GSFMODE=REPLACE
HANDSHAKE=XONXOFF
NOPROMPT
NODISPLAY
NOCELL
NOCHARACTERS
GSFNAME=XYZ;
PROC G3D;
SCATTER RTTY*RXV=V;
BY n;
RUN;
```

**A9. THE REACTION OF G TO CHANGES IN THE VALIDITY COEFFICIENT, THE CRITERION RELIABILITY COEFFICIENT AND SAMPLE SIZE, USING BOBKO & RIECK'S SECOND FORMULA FOR THE STANDARD ERROR OF THE CORRECTED CORRELATION COEFFICIENT**

```

OPTIONS LINESIZE=80 NODATE NONUMBER;
DATA DAT1;
DO RXY=.04 TO 1 BY 0.04;
DO RTTY=.1 TO .9 BY .05;
DO n=10, 90;
OUTPUT;
END;
END;
END;
DATA DAT2;
SET DAT1;
IF RXY GT SQRT(RTTY) THEN RXY=.;
RcXY=RXY/SQRT(RTTY);
SEr=(1-(RXY**2))/SQRT(n);
A=SQRT(((1/n)*(1/RTTY)*((1-RXY**2)**2)));
B=(.25/n)*(1/(RTTY**3))*(RXY**2)*((1-(RTTY**2))**2);
SErc=SQRT(A+B);
Z1=RXY/SEr;
Z2=RcXY/SErc;
G=Z2/Z1;
V=SErc/SEr;
Zk=1.6449;
IF Z1 GE Zk AND Z2 GE Zk THEN GG=1;
IF Z1 GE Zk AND Z2 LT Zk THEN GG=2;
IF Z1 LT Zk AND Z2 GE Zk THEN GG=3;
IF Z1 LT Zk AND Z2 LT Zk THEN GG=4;
IF Z1=. OR Z2=. THEN GG=.;
IF GG=1 THEN SHAPEVAR='CUBE';
IF GG=2 THEN SHAPEVAR='HEART';
IF GG=3 THEN SHAPEVAR='SPADE';
IF GG=4 THEN SHAPEVAR='FLAG';
PROC FORMAT;
VALUE GGF 1='VH01VH02'
      2='VH01AH02'
      3='AHO1VH02'
      4='AH01AH02';
PROC SORT;

```

```
BY n;  
FILENAME XYZ 'PIC9.GRAF';  
GOPTIONS  
DEVICE=HPLJGL  
AUTOFEED  
GSFMODE=REPLACE  
HANDSHAKE=XONXOFF  
NOPROMPT  
NODISPLAY  
NOCELL  
NOCHARACTERS  
GSFNAME=XYZ;  
PROC G3D;  
SCATTER RTTY*RXV=V;  
BY n;  
RUN;
```

**A10. THE REACTION OF G TO CHANGES IN THE VALIDITY COEFFICIENT, SAMPLE SIZE AND K, USING THE BOBKO & RIECK FORMULA FOR THE STANDARD ERROR OF THE CORRELATION COEFFICIENT CORRECTED FOR CASE B RESTRICTION OF RANGE**

```

OPTIONS LINESIZE=80 NODATE NONUMBER;
DATA DAT1;
DO RXY=.05 TO .95 BY 0.05;
DO K=1 TO 3 BY .1;
DO n=10, 90;
OUTPUT;
END;
END;
END;
DATA DAT2;
SET DAT1;
RcXY=(RXY*K)/(SQRT(1-(RXY**2)+(RXY**2)*(K**2)));
SEr=(1-(RXY**2))/SQRT(n);
SErc=(K/(1+(RXY**2)*((K**2)-1))**(3/2))*SEr;
Z1=RXY/SEr;
Z2=RcXY/SErc;
G=Z2/Z1;
V=SErc/SEr;
Zk=1.6449;
IF Z1 GE Zk AND Z2 GE Zk THEN GG=1;
IF Z1 GE Zk AND Z2 LT Zk THEN GG=2;
IF Z1 LT Zk AND Z2 GE Zk THEN GG=3;
IF Z1 LT Zk AND Z2 LT Zk THEN GG=4;
IF Z1=. OR Z2=. THEN GG=.;
IF GG=1 THEN SHAPEVAR='CUBE';
IF GG=2 THEN SHAPEVAR='HEART';
IF GG=3 THEN SHAPEVAR='SPADE';
IF GG=4 THEN SHAPEVAR='FLAG';
PROC FORMAT;
VALUE GGF 1='VH01VH02'
      2='VH01AH02'
      3='AHO1VH02'
      4='AH01AH02';
PROC SORT;
BY n;
FILENAME XYZ 'PIC10.GRAF';
GOPTIONS

```

```
DEVICE=HPLJGL
AUTOFEED
GSFMODE=REPLACE
HANDSHAKE=XONXOFF
NOPROMPT
NODISPLAY
NOCELL
NOCHARACTERS
GSFNAME=XYZ;
PROC G3D;
SCATTER K*RXY=V/ROTATE=10;
BY n;
RUN;
```

A11. THE REACTION OF G TO CHANGES IN THE VALIDITY COEFFICIENT, SAMPLE SIZE AND K, USING THE BOBKO & RIECK FORMULA FOR THE STANDARD ERROR OF THE CORRELATION COEFFICIENT CORRECTED FOR CASE B RESTRICTION OF RANGE AND CRITERION UNRELIABILITY [VALIDITY AND RELIABILITY CALCULATED ON SELECTED GROUP]

```

OPTIONS LINESIZE=80 NODATE NONUMBER;
DATA DAT1;
DO RXY=.05 TO .95 BY 0.05;
DO RTTY=.05 TO 1 BY .05;
DO K=5;
DO n=10, 90;
OUTPUT;
END;
END;
END;
END;
DATA DAT2;
SET DAT1;
IF RXY GT SQRT(RTTY) THEN RXY=.;
RcXY=(RXY*(RTTY**-.5)*K)/(SQRT(1-((RXY**2)/RTTY)+(((RXY**2)/RTTY)*(K**2))));
SEr=(1-(RXY**2))/SQRT(n);
D=(1-((RXY**2)*(1/RTTY))+((K**2)*(RXY**2)*(1/RTTY)));
A=(1-(RXY**2))**2;
B=.25*(RXY**2)*(1/(RTTY**2))*((1-(RTTY**2))**2);
C=(RXY)*(1/RTTY);
E=(RXY)*(1-(RXY**2)-(RTTY**2));
F=.5*(RXY)*(RTTY)*(1-2*(RXY**2)-(RTTY**2));
H=(1/n)*(1/RTTY)*((K**2)/(D**3));
SErc=SQRT(H*(A+B-(C*(E-F))));
Z1=RXY/SEr;
Z2=RcXY/SErc;
G=Z2/Z1;
V=SErc/SEr;
Zk=1.6449;
IF Z1 GE Zk AND Z2 GE Zk THEN GG=1;
IF Z1 GE Zk AND Z2 LT Zk THEN GG=2;
IF Z1 LT Zk AND Z2 GE Zk THEN GG=3;
IF Z1 LT Zk AND Z2 LT Zk THEN GG=4;
IF Z1=. OR Z2=. THEN GG=.;

```



```
IF GG=1 THEN SHAPEVAR='CUBE';
IF GG=2 THEN SHAPEVAR='HEART';
IF GG=3 THEN SHAPEVAR='SPADE';
IF GG=4 THEN SHAPEVAR='FLAG';
PROC FORMAT;
VALUE GGF 1='VH01VH02'
      2='VH01AH02'
      3='AHO1VH02'
      4='AH01AH02';
PROC PRINT;
PROC SORT;
BY n K;
FILENAME XYZ 'PIC11.GRAF';
GOPTIONS
DEVICE=HPLJGL
AUTOFEED
GSFMODE=REPLACE
HANDSHAKE=XONXOFF
NOPROMPT
NODISPLAY
NOCELL
NOCHARACTERS
GSFNAME=XYZ;
PROC G3D;
SCATTER RTTY*RXV=V;
BY n K;
RUN;
```

A12. THE REACTION OF G TO CHANGES IN THE VALIDITY COEFFICIENT, SAMPLE SIZE AND K, USING THE ALLEN & DUNBAR FORMULA FOR THE STANDARD ERROR OF THE CORRELATION COEFFICIENT CORRECTED FOR CASE C[i] RESTRICTION OF RANGE [ALL CORRELATIONS CALCULATED ON SELECTED GROUP AND USING ASSUMPTION SET A]

```

OPTIONS LINESIZE=80 NODATE NONUMBER;
DATA DAT1;
DO RXY=.04 TO .95 BY .04;
DO RXZ=.04 TO .95 BY .04;
DO RYZ=.25;
DO K=5;
DO n=10, 90;
OUTPUT;
END;
END;
END;
END;
END;
DATA DAT2;
SET DAT1;
DD1=((RXY**2)+(RXZ**2));
DD2=((RXY**2)+(RYZ**2));
DD3=((RXZ**2)+(RYZ**2));
IF DD1 GT 1 OR DD2 GT 1 OR DD3 GT 1 THEN RXY = .;
A=RXY+((K**2)-1)*RXZ*RYZ;
B=(1+((K**2)-1)*(RXZ**2));
C=(1+((K**2)-1)*(RYZ**2));
RcXY=A/(SQRT(B*C));
SEr=(1-(RXY**2))/SQRT(n);
W=(K**2)-1;
D_RXY=((1+(W*(RXZ**2)))**-.5)*((1+(W*(RYZ**2)))**-.5);
D_RXZ=W*((1+(W*(RXZ**2)))**-.5)*((1+(W*(RYZ**2)))**-.5)*(RYZ-RXY*RXZ);
D_RYZ=W*((1+(W*(RXZ**2)))**-.5)*((1+(W*(RYZ**2)))**-.5)*(RXZ-RXY*RYZ);
C_RXYRXZ=(1/n)*(RYZ*((1-(RXY**2)-(RXZ**2))**-.5)*RXY*RXZ*(1-(RYZ**2)-(RXY**2)-(RXZ**2)));
C_RXYRYZ=(1/n)*(RXZ*((1-(RXY**2)-(RYZ**2))**-.5)*RXY*RYZ*(1-(RXZ**2)-(RXY**2)-(RYZ**2)));
C_RXZRYZ=(1/n)*(RXY*((1-(RXZ**2)-(RYZ**2))**-.5)*RXZ*RYZ*(1-(RXY**2)-(RXZ**2)-(RYZ**2)));
V_RXY=((1-(RXY**2))**2)/n;
V_RXZ=((1-(RXZ**2))**2)/n;
V_RYZ=((1-(RYZ**2))**2)/n;
SErc=SQRT(((D_RXY**2)*V_RXY)+((D_RXZ**2)*V_RXZ)+((D_RYZ**2)*V_RYZ)+
(2*D_RXY*D_RXZ*C_RXYRXZ)+(2*D_RXY*D_RYZ*C_RXYRYZ)+(2*D_RXZ*D_RYZ*C_RXZRYZ));

```

```
Z1=RXY/SEr;
Z2=RcXY/SErc;
G=Z2/Z1;
V=SErc/SEr;
Zk=1.6449;
IF Z1 GE Zk AND Z2 GE Zk THEN GG=1;
IF Z1 GE Zk AND Z2 LT Zk THEN GG=2;
IF Z1 LT Zk AND Z2 GE Zk THEN GG=3;
IF Z1 LT Zk AND Z2 LT Zk THEN GG=4;
IF Z1=. OR Z2=. THEN GG=;
IF GG=1 THEN SHAPEVAR='CUBE';
IF GG=2 THEN SHAPEVAR='HEART';
IF GG=3 THEN SHAPEVAR='SPADE';
IF GG=4 THEN SHAPEVAR='FLAG';
PROC FORMAT;
VALUE GGF 1='VH01VH02'
      2='VH01AH02'
      3='AHO1VH02'
      4='AH01AH02';
PROC SORT;
BY n K RYZ;
FILENAME XYZ 'PIC12.GRAF';
GOPTIONS
DEVICE=HPLJGL
AUTOFEED
GSFMODE=REPLACE
HANDSHAKE=XONXOFF
NOPROMPT
NODISPLAY
NOCELL
NOCHARACTERS
GSFNAME=XYZ;
PROC G3D;
SCATTER RXY*RXZ=SErc/ROTATE=10;
BY n K RYZ;
RUN;
```

**A13. THE REACTION OF J TO CHANGES IN THE VALIDITY COEFFICIENT, THE CRITERION RELIABILITY COEFFICIENT AND SAMPLE SIZE, USING BOBKO & RIECK'S FIRST FORMULA FOR THE STANDARD ERROR OF THE CORRECTED CORRELATION COEFFICIENT**

```

OPTIONS LINESIZE=80 NODATE NONUMBER;
DATA DAT1;
DO RHOxy=.04 TO 1 BY 0.04;
DO RHOttY=.1 TO .9 BY .05;
DO n=10, 90;
OUTPUT;
END;
END;
END;
DATA DAT2;
SET DAT1;
IF RHOxy GT SQRT(RHOttY) THEN RHOxy=.;
RcXY=RHOxy/SQRT(RHOttY);
SEr=(1-(RHOxy**2))/SQRT(n);
A=SQRT((1/n)*(1/RHOttY)*((1-RHOxy**2)**2));
B=(.25/n)*(1/(RHOttY**3))*(RHOxy**2)*((1-(RHOttY**2))**2);
C=((1/n)*(1/(RHOttY**2))*(RHOxy))*((RHOttY)*(1-(2*(RHOxy**2)))-.5*(RHOxy**2)*(1-(2*(RHOxy**2))-(RHOttY**2))));
SErc=SQRT(A+B-C);
Z1=RHOxy/SEr;
Z2=RcXY/SErc;
J=Z1/Z2;
Zk=1.6449;
PROC SORT;
BY n;
FILENAME XYZ 'PIC13.GRAF';
GOPTIONS
DEVICE=HPLJGL
AUTOFEED
GSFMODE=REPLACE
HANDSHAKE=XONXOFF
NOPROMPT
NODISPLAY
NOCELL
NOCHARACTERS
GSFNAME=XYZ;
PROC G3D; SCATTER RHOttY*RHOxy=J; BY n; RUN;

```

A14. THE REACTION OF J TO CHANGES IN THE VALIDITY COEFFICIENT, SAMPLE SIZE AND K, USING THE ALLEN & DUNBAR FORMULA FOR THE STANDARD ERROR OF THE CORRELATION COEFFICIENT CORRECTED FOR CASE C[i] RESTRICTION OF RANGE [ALL CORRELATIONS CALCULATED ON SELECTED GROUP AND USING ASSUMPTION SET A]

```

OPTIONS LINESIZE=80 NODATE NONUMBER;
DATA DAT1;
DO RXY=.04 TO .95 BY .04;
DO RXZ=.04 TO .95 BY .04;
DO RYZ=.15, .65;
DO K=2;
DO n=90;
OUTPUT;
END;
END;
END;
END;
END;
DATA DAT2;
SET DAT1;
DD1=((RXY**2)+(RXZ**2));
DD2=((RXY**2)+(RYZ**2));
DD3=((RXZ**2)+(RYZ**2));
IF DD1 GT 1 OR DD2 GT 1 OR DD3 GT 1 THEN RXY = .;
A=RXY+((K**2)-1)*RXZ*RYZ;
B=(1+((K**2)-1)*(RXZ**2));
C=(1+((K**2)-1)*(RYZ**2));
RcXY=A/(SQRT(B*C));
SEr=(1-(RXY**2))/SQRT(n);
W=(K**2)-1;
D_RXY=((1+(W*(RXZ**2)))**-.5)*((1+(W*(RYZ**2)))**-.5);
D_RXZ=W*((1+(W*(RXZ**2)))**-.5)*((1+(W*(RYZ**2)))**-.5)*(RYZ-RXY*RXZ);
D_RYZ=W*((1+(W*(RXZ**2)))**-.5)*((1+(W*(RYZ**2)))**-.5)*(RXZ-RXY*RYZ);
C_RXYRXZ=(1/n)*(RYZ*((1-(RXY**2)-(RXZ**2))**-.5))*RXY*RXZ*(1-(RYZ**2)-(RXY**2)-(RXZ**2));
C_RXYRYZ=(1/n)*(RXZ*((1-(RXY**2)-(RYZ**2))**-.5))*RXY*RYZ*(1-(RXZ**2)-(RXY**2)-(RYZ**2));
C_RXZRYZ=(1/n)*(RXY*((1-(RXZ**2)-(RYZ**2))**-.5))*RXZ*RYZ*(1-(RXY**2)-(RXZ**2)-(RYZ**2));
V_RXY=((1-(RXY**2))**2)/n;
V_RXZ=((1-(RXZ**2))**2)/n;
V_RYZ=((1-(RYZ**2))**2)/n;
SErc=SQRT(((D_RXY**2)*V_RXY)+((D_RXZ**2)*V_RXZ)+((D_RYZ**2)*V_RYZ)+
(2*D_RXY*D_RXZ*C_RXYRXZ)+(2*D_RXY*D_RYZ*C_RXYRYZ)+(2*D_RXZ*D_RYZ*C_RXZRYZ));
Z1=RXY/SEr;

```

```
Z2=RcXY/SErc;  
J=Z1/Z2;  
Zk=1.6449;  
IF Z1 GE Zk AND Z2 GE Zk THEN GG=1;  
IF Z1 GE Zk AND Z2 LT Zk THEN GG=2;  
IF Z1 LT Zk AND Z2 GE Zk THEN GG=3;  
IF Z1 LT Zk AND Z2 LT Zk THEN GG=4;  
PROC SORT;  
BY n K RYZ;  
FILENAME XYZ 'PIC14.GRAF';  
GOPTIONS  
DEVICE=HPLJGL  
AUTOFEED  
GSFMODE=REPLACE  
HANDSHAKE=XONXOFF  
NOPROMPT  
NODISPLAY  
NOCELL  
NOCHARACTERS  
GSFNAME=XYZ;  
PROC G3D;  
SCATTER RXY*RXZ=J;  
BY n K RYZ;  
RUN;
```



A15. THE REACTION OF J TO CHANGES IN THE VALIDITY COEFFICIENT, SAMPLE SIZE AND K, USING THE BOBKO & RIECK FORMULA FOR THE STANDARD ERROR OF THE CORRELATION COEFFICIENT CORRECTED FOR CASE B RESTRICTION OF RANGE AND CRITERION UNRELIABILITY [VALIDITY AND RELIABILITY CALCULATED ON SELECTED GROUP]

```

OPTIONS LINESIZE=80 NODATE NONUMBER;
DATA DAT1;
DO RXY=.05 TO .95 BY 0.05;
DO RTTY=.05 TO 1 BY .05;
DO K=2;
DO n=10, 90;
OUTPUT;
END;
END;
END;
END;
DATA DAT2;
SET DAT1;
IF RXY GT SQRT(RTTY) THEN RXY=.;
RcXY=(RXY*(RTTY**-.5)*K)/(SQRT(1-((RXY**2)/RTTY)+(((RXY**2)/RTTY)*(K**2))));
SEr=(1-(RXY**2))/SQRT(n);
D=(1-((RXY**2)*(1/RTTY))+((K**2)*(RXY**2)*(1/RTTY)));
A=(1-(RXY**2))**2;
B=.25*(RXY**2)*(1/(RTTY**2))*((1-(RTTY**2))**2);
C=(RXY)*(1/RTTY);
E=(RXY)*(1-(RXY**2)-(RTTY**2));
F=.5*(RXY)*(RTTY)*(1-2*(RXY**2)-(RTTY**2));
H=(1/n)*(1/RTTY)*((K**2)/(D**3));
SErc=SQRT(H*(A+B-(C*(E-F))));
Z1=RXY/SEr;
Z2=RcXY/SErc;
J=Z1/Z2;
PROC SORT;
BY n K;
FILENAME XYZ 'PIC15.GRAF';
GOPTIONS
DEVICE=HPLJGL
AUTOFEED
GSFMODE=REPLACE
HANDSHAKE=XONXOFF
NOPROMPT

```

```
NODISPLAY  
NOCELL  
NOCHARACTERS  
GSFNAME=XYZ;  
PROC G3D;  
SCATTER RTTY*RXY=J/ROTATE=40;  
BY n K;  
RUN;
```

**A16. THE REACTION OF W TO CHANGES IN Z[X] FOR DIFFERENT SCENARIOS DEFINED IN TERMS OF THE VALIDITY COEFFICIENT AND THE PREDICTOR AND CRITERION RELIABILITY COEFFICIENTS**

```

OPTIONS LINESIZE=80;
DATA DAT1;
DO ZX = .1 TO 6 BY .2;
DO R = .3, .6;
DO RTTY = .6, .8, .9;
DO RTTX = .5, .6, .8;
OUTPUT;
END;
END;
END;
END;
DATA DAT2;
SET DAT1;
ZX=-1*ZX;
ZY=1.5;
Q=(SQRT(RTTY))*(SQRT(RTTX));
IF R GT Q THEN R=.;
A=(R*ZX)*((SQRT(RTTX))-1);
B=(SQRT(1-(R**2)))-(SQRT(RTTY-(R**2)));
DELTA=A/B;
ZC=(ZY-(R*ZX))/(SQRT(1-(R**2)));
W=DELTA*(1/(ZC));
DATA DAT3;
SET DAT2;
PROC SORT;
BY R RTTX RTTY;
PROC PRINT;
VAR ZX W;
BY R RTTX RTTY;
RUN;

```

A17. THE REACTION OF THE RATIO  $\beta[Y|X]/\beta[y|x]$  TO CHANGES IN  $K$ ,  $\rho[x,y]$ ,  $\rho[x,z]$  AND  $\rho[y,z]$ 

```

OPTIONS LINESIZE=80;
DATA DAT1;
DO RXY=.1;
DO RXZ=.1 TO .9 BY .05;
DO RYZ=.1 TO .9 BY .05;
DO K=2, 4;
OUTPUT;
END;
END;
END;
END;
DATA DAT2;
SET DAT1;
R=SQRT(((RXZ**2)+(RYZ**2)-(2*RXZ*RYZ*RXY))/(1-(RXY**2)));
IF R GT 1 THEN RXY=.;
B_YXR=RXY;
A=RXY-(RYZ*RXZ)+(RYZ*RXZ)*(K**2);
B=1-(RXZ**2)+(RXZ**2)*(K**2);
B_YXU=A/B;
DELTA=B_YXU/B_YXR;
PROC SORT;
BY RXY K;
DATA DAT3;
SET DAT2;
IF RXY NE .;
FILENAME XYZ 'PIC19.GRAF';
GOPTIONS
DEVICE=HPLJGL
AUTOFEED
GSFMODE=REPLACE
HANDSHAKE=XONXOFF
NOPROMPT
NODISPLAY
NOCELL
NOCHARACTERS
GSFNAME=XYZ;
PROC G3D;
SCATTER RYZ*RXZ=DELTA;
BY RXY K;
RUN;

```

A18. THE REACTION OF THE RATIO  $\sigma[Y|X]/\sigma[y|x]$  UNDER CASE C[i] RESTRICTION OF RANGE  
TO CHANGES IN K,  $\rho[x,y]$ ,  $\rho[x,z]$  AND  $\rho[y,z]$

```
OPTIONS LINESIZE=80;
DATA DAT1;
DO ZY = -3 TO 3 BY .05;
DO RTTY = .1 TO .9 BY .05;
OUTPUT;
END;
END;
DATA DAT2;
SET DAT1;
BR = (1-PROBNORM(ZY));
ZYC = (ZY/SQRT(RTTY));
BRC = (1-PROBNORM(ZYC));
DELTA=BR-BRC;
FILENAME XYZ 'PIC20.GRAF';
GOPTIONS
DEVICE=HPLJGL
AUTOFEED
GSFMODE=REPLACE
HANDSHAKE=XONXOFF
NOPROMPT
NODISPLAY
NOCELL
NOCHARACTERS
GSFNAME=XYZ;
PROC G3D;
PLOT BR*RTTY=DELTA/ROTATE=0;
RUN;
```

ξψζξψζξψζξψζξψζξψζξψζξψζξψζξψζ